Plant Molecular Breeding

Biological Sciences Series

A series which provides an accessible source of information at research and professional level in chosen sectors of the biological sciences.

Series Editors:

Professor Jeremy A. Roberts, Plant Science Division, School of Biosciences, University of Nottingham.
Professor Peter N. R. Usherwood, Molecular Toxicology Research Group, School of Life and Environmental Sciences, University of Nottingham.

Titles in the series:

Stress Physiology in Animals
Edited by P. H. M. Balm

Seed Technology and its Biological Basis
Edited by M. Black and J. D. Bewley

Leaf Development and Canopy Growth
Edited by B. Marshall and J. A. Roberts

Environmental Impacts of Aquaculture
Edited by K. D. Black

Herbicides and their Mechanisms of Action
Edited by A. H. Cobb and R. C. Kirkwood

The Plant Cell Cycle and its Interfaces
Edited by D. Francis

Meristematic Tissues in Plant Growth and Development
Edited by M. T. McManus and B. E. Veit

Fruit Quality and its Biological Basis
Edited by M. Knee

Pectins and their Manipulation
Edited by G. B. Seymour and J. P Knox

Wood Quality and its Biological Basis
Edited by J. R. Barnett and G. Jeronimidis

Plant Molecular Breeding
Edited by H. J. Newbury

Biogeochemistry of Marine Systems
Edited by K. D. Black and G. Shimmield

Plant Molecular Breeding

Edited by

H. JOHN NEWBURY
School of Biosciences
University of Birmingham
Edgbaston, Birmingham
UK

Blackwell
Publishing

CRC Press

Editorial offices:
Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK
Tel: +44 (0)1865 776868
Blackwell Publishing Asia Pty Ltd, 550 Swanston Street, Carlton, Victoria 3053, Australia
Tel: +61 (0)3 8359 1011

ISBN 1-84127-321-X
Originated as Sheffield Academic Press

Published in the USA and Canada (only) by CRC Press LLC, 2000 Corporate Blvd., N.W., Boca Raton, FL 33431, USA
Orders from the USA and Canada (only) to CRC Press LLC

USA and Canada only:
ISBN 0-8493-2813-X

First published 2003

Library of Congress Cataloging-in-Publication Data:
A catalog record for this title is available from the Library of Congress

British Library Cataloguing-in-Publication data:
A catalogue record for this title is available from the British Library

Set in 10/12 pt Times
by Sparks Computer Solutions Ltd
http://www.sparks.co.uk
Printed and bound in Great Britain using acid-free paper by MPG Books Ltd, Bodmin, Cornwall

For further information on Blackwell Publishing, visit our website:
www.blackwellpublishing.com

Contents

Contributors

James A. Anderson	Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, Minnesota, USA
M. W. Bonierbale	Centro Internacional de la Papa (CIP), Apartado 1558, Lima 12, Peru
Keith J. Edwards	School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK
M. Ghislain	Centro Internacional de la Papa (CIP), Apartado 1558, Lima 12, Peru
Ian Godwin	School of Land and Food Sciences, University of Queensland, Brisbane QLD 4072, Australia
Frédéric Hospital	INRA, Station de Génétique Végétale, Ferme du Moulon, 91190 Gif sur Yvette, France
Michael J. Kearsey	School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK
X-Q. Li	Potato Research Centre, Agriculture and Agri-Food Canada, PO Box 20280, Fredericton, New Brunswick EB3 4Z7, Canada
Zewei W. Luo	School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK

C. Mba — Centro Internacional de Agricultura Tropical (CIAT), A. A. 6713, Cali, Colombia

H. John Newbury — School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK

Andy H. Paterson — College of Agricultural and Environmental Sciences, University of Georgia, Athens, GA 30602, USA

Ian Puddephat — Horticulture Research International, Wellesbourne, Warwickshire, CV35 9EF, UK
Now at: Syngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, UK

R. Simon — Centro Internacional de la Papa (CIP), Apartado 1558, Lima 12, Peru

Donal M. O'Sullivan — Project Leader, Molecular Research Group, NIAB, Huntingdon Road, Cambridge CB3 0LE, UK

D. P. Zhang — Centro Internacional de la Papa (CIP), Apartado 1558, Lima 12, Peru

Preface

The past few years have seen an explosion of new information and resources in the area of plant molecular genetics and genomics. As a result of developments in high-throughput sequencing and user-friendly databases with easy access via the internet, we now have available huge amounts of information on plant genes and physical and genetic maps. The milestone whole genome sequencing of both *Arabidopsis* and rice provides the most obvious evidence of progress in this area. For these model species, we have entered the era of functional genomics, which aims to determine the functions of all the genes identified in their genomes. A community-wide effort is underpinning studies in this area, and the insights provided by these analyses make this an exciting time to be a plant biologist. But how does the growing mountain of information on gene structure, organization and function help people charged with the task of improving crop species? This is one of the central themes of this book, in which researchers from leading laboratories around the world provide insights into their specialized areas to provide an overview of the state of molecular plant breeding at the beginning of the twenty-first century.

It is usual to classify attempts to improve plants into one of two strategies: by plant breeding, or by plant genetic engineering. In fact, improvement programmes frequently use a combination of these approaches. In any case, a thorough understanding of both classical and molecular genetics is required for either technique to be successful and efficient. Most of the traits that are the subject of improvement programmes are quantitative in nature. The chapter on the mapping, characterization and deployment of quantitative genetic trait loci (QTL) introduces the reader to the convergence between the statistical approaches of quantitative genetics and genomic sequence data that together allow the proposal of candidate genes for QTL controlling important crop traits. The chapter on genomic colinearity describes how an understanding of the conserved organization of genes between plant taxa facilitates the exploitation of genetic information derived from model species in less-heavily studied crops. Exploitation of wild relatives of crop species has also been reviewed, in order to explain the value of plant genetic resources as a supply of novel alleles for crop improvement. All three chapters demonstrate the importance of an array of molecular marker technologies for the location and characterization of plant genes.

In many cases, genes identified using the approaches just described can be deployed in breeding programmes, and the chapter describing marker-assisted breeding analyses the efficiency of different marker-based methods for following the inheritance of target genes in progeny plants. In other cases, isolated genes can be

transferred using genetically modified (GM) technology, and an extensive review of recent advances in plant genetic engineering has been included. Finally, three crop types have been selected for a detailed analysis of the effects of developments in molecular genetics on improvement programmes. The choice of wheat and maize as examples is based upon their relative global importance as crop plants. The chapter on root and tuber crop improvement has a more commodity-based perspective and offers the opportunity to compare the advantages and disadvantages of working with several starch crops.

H.J. Newbury

1 Mapping, characterization and deployment of quantitative trait loci

Michael J. Kearsey and Zewei W. Luo

1.1 Introduction

Most of the very considerable progress in genetics over the last century has focused on using single gene mutants that produce relatively clear-cut effects on the phenotype. They have involved comparisons between normal and disfunctional alleles, so effects are very large compared with the background variation at other gene loci and the environment. As a result, their inheritance can easily be seen to follow Mendelian laws. However, most natural variation of importance to plant breeders (yield, emergence time, stress tolerance, etc.) is not of this sort. The phenotypes in an F_2 do not fall into clear-cut Mendelian ratios, but most commonly show a continuous, approximately normal, range of variation. The analysis and deployment of quantitative trait loci is therefore of enormous importance in breeding programmes. Although it is possible to undertake breeding programmes using only phenotypic selection, an understanding of the number and location of quantitative trait loci (QTL) controlling performance for a target trait can markedly enhance the efficiency of breeding. In this chapter, we will be reviewing the theoretical background to quantitative trait analysis and explaining how the concepts and statistical approaches that arise from these theoretical considerations can be employed to analyse the genetic basis of quantitative traits in plants. Most of the progress in QTL analysis over the past decade has occurred because of the availability of a range of informative molecular marker techniques. We will explain the methods by which associations between the inheritance of alleles at marker and trait loci allow the identification and mapping of QTL.

Information about QTL location, effects and even the sequence can be exploited in a number of ways in crop improvement programmes, and these are explained in later chapters in this book. In Chapter 2, the reader is shown how map location information can be used in a range of marker-assisted selection protocols. In Chapter 3, there is a review of the value of synteny between plant genomes. It is clear from the information presented that knowledge of the map location of an important QTL in one species allows the accurate prediction of the map location of that QTL in related species. This means, for example, that knowledge of the location of an important flowering time QTL in rice allows one to target specific chromosome intervals in less well-studied cereals in breeding programmes concerned with earliness. Re-

cent advances in plant molecular genetics have eventually allowed the cloning of QTL largely because of the information produced by the sequencing of the entire genome of the model species *Arabidopsis.* Hence, QTL – which were once largely theoretical genetic determinants that helped explain trait performance in statistical analyses – are now being sequenced and the allelic variations that account for their different effects are being elucidated. The isolation of QTL as cloned DNA fragments opens the possibility of the transfer of QTL alleles by genetic engineering techniques and GM technologies are reviewed in Chapter 4.

1.2 Genetic basis of quantitative trait performance

It has been generally assumed, ever since the pioneering work of Nilsson-Ehle (1909) almost 100 years ago, that continuous variation in trait performance is due to the joint segregation of several genes, all of which have a small but quasi-additive effect on the phenotype, together with a major effect of the environment. It has been surmised that the allelic differences for these genes are small. This is because all alleles are presumed to be functional but have slightly different efficiencies in their contribution to the trait in question. The genes responsible for such traits were originally called polygenes by Mather (1941), but are now generally referred to as QTL (Gelderman 1975). Their existence was indicated from the fact that selection for five or ten generations from an F_2 population normally resulted in extreme lines that transgressed the range of the original population. This indicated that the allelic variation was already present in the F_2 because the time scale was too short and the responses too repeatable for new mutations to have been responsible. Also, the transgressive segregation indicated that new genotypes had been formed that were not present in the F_2 and this was only possible if two or more genes were involved. Their existence was confirmed by elegant experiments by Breese and Mather (1957, 1960) and Thoday (1961, 1979) using fruit flies. Their approach was to construct breeding lines that contained various combinations of chromosomal segments from two different parental lines. They showed that several regions of the chromosome could be associated with any given trait and that these segments did, indeed, contribute almost additively. However, at that time, they only had access to major gene mutants to identify and manipulate the segments of chromosome during the construction of the lines used to analyse their material, and these mutants sometimes also affected the traits concerned. It was not a feasible approach to be easily adapted to plant breeding. However, several workers have capitalized on the use of aneuploids in wheat to achieve the location of QTL first to whole chromosomes and then to parts of chromosomes in wheat (Sears 1953; Law 1967; Law *et al.* 1983).

Because of the difficulties of identifying the individual QTL, little progress was made in actually studying their detailed nature until the advent of molecular, DNA markers in the late 1980s (Lander & Botstein 1989). Thus, it was not – and still is

not – generally known whether QTL are structural or regulatory genes, and we have no clear idea of how many there are nor the precise nature of the allelic variation. Nonetheless, quantitative geneticists have been very successful in obtaining useful information to predict response to selection, understand heterosis, optimize breeding strategies and obtain general information on the type of gene action and interaction underlying the traits (Falconer & Mackay 1996; Kearsey & Pooni 1996; Lynch & Walsh 1998). This has all been achieved simply by studying the correlations in phenotypes between relatives in a range of family structures. Quantitative geneticists adopted the view that it was easier to assume that there were many QTL of roughly equal effect and to manipulate their combined effects rather than to try and dissect their individual components. Even if the individual components were known, it was argued, it would still be difficult accurately to predict how they might work in combination.

1.3 Basic modelling of quantitative traits

It was assumed that the individual QTL follow all the basic laws of Mendelian inheritance. Thus, they segregate independently at meiosis, sometimes exhibit linkage to other QTL, show some degree of dominance though probably not over-dominance, and they could show gene interaction. Based on this basic, yet simple structure, it is possible to construct models to explain the means and variances of various types of family. The essential principles were set down by Fisher (1918) but have been considerably developed subsequently (Mather & Jinks 1982; Falconer & Mackay 1996; Kearsey & Pooni 1996; Lynch & Walsh 1998).

Consider an F_2 derived from two parental inbred lines that were identical at all genes except one, gene A. If we call the two alleles of gene A, A^+ and A^-, and the + and – indicate whether the allele increases or decreases the trait, then A will segregate in the F_2 with frequencies and genetic values as shown below:

F_2 genotypes	A^-A^-	A^+A^-	A^+A^+	mean
Frequency	¼	½	¼	
Genetic value	$m-a$	$m+d$	$m+a$	$m+½d$

It is customary to define parameters such as a and d as deviations from the mean of the two homozygotes m, although there are other approaches. The genetic value is the model of the parameters contributing to the mean phenotype of each genotype. Thus, replacing the one homozygous allele with the other causes the mean to change by $2a$. The homozygous effect, a is referred to as the additive genetic effect as opposed to d the dominance effect. The value of d is assumed to be between +/– a in size; that is, an allele may show zero, partial or complete dominance but, based on most known alleles for major genes, over-dominance, that is where $d > a$, is **not** expected to occur.

The genetic variation among the F_2 individuals around the F_2 mean for this single gene is:

$$\Sigma(x - \bar{x})^2 = \tfrac{1}{4}(-a - \tfrac{1}{2}d)^2 + \tfrac{1}{2}(d - \tfrac{1}{2}d)^2 + \tfrac{1}{4}(a - \tfrac{1}{2}d)^2$$

$$= \tfrac{1}{4}(-a)^2 + \tfrac{1}{2}(d)^2 + \tfrac{1}{4}(a)^2 - (\tfrac{1}{2}d)^2$$

$$= \tfrac{1}{2}a^2 + \tfrac{1}{4}d^2$$

This simply states that, for a single gene, A, the genetic variation between individuals in an F_2 population will consist of the squared additive and dominance effects in the ratio ½: ¼. This will be true for all other genes providing that they are independent in action and inheritance; that is, they do not interact and they are not linked. If these assumptions hold, then the combined variance of all the genes controlling the trait is simply the sum of their individual components. Thus the combined genetic variance, V_G will be:

$$V_G = \tfrac{1}{2}\Sigma a^2 + \tfrac{1}{4}\Sigma d^2 = V_A + V_D \qquad (1.1)$$

where V_A and V_D are the variances due to additive and dominance variation respectively. The overall phenotypic variation (V_P) will also include environmental variation (V_E) and so

$$V_P = V_A + V_D + V_E \qquad (1.2)$$

This is a fundamental equation of quantitative genetics and applies to all populations not just to F_2s, although the exact formulations of V_A and V_D will depend on the nature of the population being studied.

The relative contributions of genetic and additive genetic variation to a particular trait in a particular environment are referred to as the broad (h^2_b) and narrow (h^2_n) heritabilities of the trait, respectively. The former ($h^2_b = V_G/V_P$) indicates the overall contribution of genetical variation to the trait and so puts a limit on the extent to which the combined effects of individual genes can contribute as we will see later. The latter ($h^2_n = V_A/V_P$) indicates the amount of variation available for selection and can be used to predict the response to selection.

Equation 1.1 above indicates that a few large genes can have a disproportionate effect on V_G because it is their squared effects that are important. Thus, two genes with additive effects, *a*, of 3 and 1 units each will contribute 9 and 1 units to the variance. As we will see later, when the individual QTL are located and their effects, *a*, estimated, it is normally their relative contribution to the variance that is quoted. Of course, to a plant breeder, it is their relative contribution to the mean that may be most relevant because varieties are sold on the basis of their means, not their variances. However, if those genes with the greatest effect on a particular trait could be

located and selection focused on them alone, it would provide an obvious benefit to breeders.

In what follows we discuss how the individual QTL can be located and their additive (a) and dominance (d) effects estimated. This will lead into how they can be manipulated in breeding programmes and, if located accurately enough, cloned.

1.4 Statistical principles and methods for mapping QTL

In order to obtain a better understanding of quantitative genetic, polygenic variation it is necessary to ask fundamental questions about the number, genomic positions, genetic effects and interactions of quantitative trait loci, QTL (Mather & Jinks 1982). The central approach to QTL location and analysis in the era of structural genomics is to attempt to correlate the genetic variation in a given quantitative trait with polymorphic genomic regions identified by molecular markers. This is an essential primary step for the ultimate identification of candidate genes and also to pursue our understanding of the molecular basis underlying the variation.

The basic idea behind QTL mapping is no more than that for mapping genes controlling morphological traits that show a simple pattern of Mendelian segregation, as in classical linkage studies. The degree of co-segregation of genes at different loci reflects the genetic distance between the loci under question, but the co-segregation has to be modelled and analysed using different approaches for simple Mendelian traits and quantitative traits. For the former, the pattern of phenotypic variation provides full information about the genotypic segregation at the loci because there is a one-to-one correspondence between phenotype and genotype. Thus, the gene-mapping problem can be based entirely on surveying the frequency of recombinant genotypes observed from experimental trials. For quantitative traits, on the other hand, such a one-to-one relationship no longer exists because the phenotype is the result of several genes and the environment. Therefore, the phenotypic variation provides only partial information about the segregation of the underlying genes, so the key problem for mapping quantitative trait loci is to uncover genotypic information about each individual QTL from relevant marker mapping data, using appropriate statistical methods.

Data for mapping QTL consist of three resources: trait phenotype; polymorphic genetic markers; and genetic structure of mapping populations. The phenotypic record of an individual for a trait reflects the genetic effects of QTL alleles that the individual carries as well as environmental contributions to the development of the character. Because markers have individually recognizable effects, they can be tracked and mapped like major genes. The genetic structure of a mapping population defines the domain in which genes at individual QTL segregate and the pattern of recombination between genes at linked loci. The statistical task of QTL mapping analysis is essentially to bridge the relationship between the trait phenotype with the genotype at the genomic regions specified by the marker loci. In this section, we review the development of the major statistical tools used in QTL analysis and ex-

plore their properties and utilities in analysing mapping experiments. Because these methods were developed for different mapping populations, the following discussion is organized on the basis of population type.

1.4.1 Molecular markers for QTL mapping

QTL can only be mapped by following their co-segregation with other markers, and it has been the proliferation of simple, reliable molecular marker methods that has been responsible for much of the progress in this area over recent years. There is not space here to present a full review of molecular marker technologies, but some basic points will be made. For a more detailed coverage of this area, the reader is referred to Staub *et al.* (1996) and Westman and Kresovich (1997).

In the approximate chronological order of their development, the major molecular marker types have included isoenzymes, RFLPs, SSRs, AFLPs and SNPs (see below). Isoenzyme methods depend upon the electrophoretic separation of proteins in a non-denaturing gel. This is followed by enzyme-specific staining which allows the visualization of bands of coloured reaction products (Hamrick & Godt 1990). The technique is robust and the marker is co-dominant (both alleles can be scored in a heterozygote), but the number of markers that can be employed is severely limited by the number of enzyme-specific stains that are available.

The scoring of restriction fragment length polymorphisms (RFLPs) requires the availability of sets of DNA sequences that can be used as (normally radioactive) probes (Tanksley *et al.* 1989). These are often cDNAs, selected from a library produced for the species under study. They are used to hybridize to homologous sequences on DNA fragments that have been produced by the digestion of the genome of test genotypes by a restriction enzyme (such as *Eco*RI). Genomic fragments, which may vary in size in different genotypes, are separated by gel electrophoresis and then transferred to a membrane filter by blotting before probe hybridization. Again, the technique is robust and the marker is co-dominant. RFLP technology remains extremely useful in genetic mapping, but in recent years has often been replaced by polymerase chain reaction (PCR)-based methods that require smaller quantities of DNA (and hence tissue for extraction) and are usually faster.

Simple sequence repeats (SSRs) are a widely used marker type that relies upon the high rate of polymorphism observed at microsatellite loci (Goldstein & Sclotterer 1999; Morgante & Olivieri 1993). These are tandem repeats of short units (usually one to four bases) that are widespread within eukaryotic genomes. Variation in the numbers of repeats is observed by developing locus-specific primers that anneal to sequences flanking the repeat region, and then using PCR to amplify the intervening DNA fragment. Alleles are visualized as bands with differing mobilities on a gel, with the marker again being co-dominant in nature. The hyper-variability in the microsatellite repeat numbers means that one is very likely to detect different alleles when one genotypes two parents used in a cross. However, considerable work has to be carried out in a species to obtain sequence data at exploitable loci before routine genotyping can be achieved.

Amplified fragment length polymorphisms (AFLPs), on the other hand, require no such preliminary work (Vos *et al.* 1997). The same kits of oligonucleotides can be used with any plant species. In this technique, DNA is digested with two different restriction enzymes (e.g. *Eco*R1 and *Mse*1) creating differing 'sticky ends'. Different adapters (short double-stranded DNA sequences) are added to the different sticky ends, after which primers specific to the two adapters are used to direct amplification of the fragments. As described, this would lead to the amplification of every restriction fragment leading to an uninformative smear of bands on a gel. However, the critical characteristic of the AFLP technique is that the primers used for amplification carry short extensions (or 'anchors', that are typically three bases long) at their 3' ends so that only a small sub-set of adapter-ligated restriction fragments is selectively amplified. One of the primers used is usually radioactively labelled, and the amplification products are typically separated on a large polyacrylamide sequencing gel. This is dried and 50 to 100 bands can normally be scored when it is subjected to autoradiography. The large numbers of markers obtained are offset by the disadvantages that the technique is more complex than most other PCR-based methods and that the markers are dominant: the scored alleles are 'band present' and 'band absent'. Increasingly, automated DNA sequencers – which are in essence DNA fragment analysers – are being used for the separation and scoring of amplification products produced by the SSR and AFLP procedures. Automated DNA sequencers require the use of fluorescent labels for fragment detection. By attaching different fluorescent labels to individual primers, one can distinguish amplification fragments obtained for different loci in the same sample. This so-called 'multiplexing' – along with the fact that many sequencers can analyse several 96-well plates of samples in a few hours and that liquid-handling robots can be used to set up the necessary PCRs – has allowed many commercial breeding programmes to make use of high-throughput genotyping.

The most recent molecular marker method to be used for plant genotyping is single nucleotide polymorphism (SNP) technology (Schafer & Hawkins 1998; Chicurel 2001). This name is somewhat misleading, since the difference between alleles at polymorphic RFLP or AFLP loci may be due to single nucleotide changes. However, SNP technology has become the generic term used for a series of high-throughput methods that each directly scores alleles that differ by a single point mutation at specific loci. The methods by which differing alleles are scored vary considerably. In some cases this involves the use of a sequencer following a primer extension protocol, while in other cases it requires the use of mass spectrometry to distinguish allelic DNA fragments on the basis of their exact mass. In either case, routine application of SNP technology for molecular marker studies requires that the DNA sequences of sets of loci are known for both parents of a cross. For the current user in an academic environment, this largely restricts genome-wide SNP genotyping to studies of *Arabidopsis* for which SNPs between the Columbia and Landsberg erecta genotypes have been made publicly available. However, in a commercial context, where extensive sequencing data may be held on particular crop species, this ap-

proach offers the opportunity for rapid and reliable genotyping of large numbers of progeny plants.

1.4.2 QTL mapping in segregating populations

The most commonly used populations for QTL analysis in crop and some model animal species are segregating populations created from two inbred lines or strains. These strains are usually assumed to be homozygous with different alleles at both QTL and genetic markers. These refer mainly to F_2 and backcross populations but also to their inbred derivatives (see section 1.5.1). There are several major advantages of using such segregating populations for QTL mapping analysis:

1. Simple bi-allelic segregation at each of these loci.
2. Full information about the linkage phase of genes at the marker loci and QTL.
3. Ease in creating a large full-sib family size.
4. Versatility of the experimental design for both detecting marker-QTL linkage and estimating genetic parameters defining genetic effects at the QTL.

The most powerful statistical method for modelling segregating populations for QTL mapping can be traced back to the benchmark paper by Lander and Botstein (1989). The method has come to be known as 'interval mapping' because it systematically searches all possible QTL locations within every chromosomal interval flanked by a pair of adjacent marker loci. Interval mapping analysis considers the following linear model to test for and to localize a putative QTL on an interval flanked by markers M_i and M_{i+1}. Under the model, the phenotypic record (y_j) of the j^{th} individual within a random sample of size n from a segregating population is given as

$$y_j = u + u_k + e_j$$

where: u is the population mean; u_k is the genetic value of the genotype of that QTL where k represents the particular QTL genotype (of which there are three possible in an F_2); e_j reflects the residual random variation in the model which is assumed to follow a normal distribution with mean zero and the variance σ^2.

The difficulty of statistical inference based on the above model lies in the fact that the genotype at the QTL is unknown, and so its genetic effect, u_k will have to be predicted from analysing the likelihood given below. The formula states the likelihood of there being a QTL at a particular location, its genetic effect and residual variance due to environmental variation and genetic segregation at other QTL, given the data of trait phenotype and genotype at the flanking marker loci. Statistical analysis of QTL mapping essentially involves a search of the likelihood function for values of the parameters which maximize the likelihood.

$$L_q(u;u_1,u_2,..u_S;\sigma^2) = \prod_{j=1}^{n} \{\sum_{k=1}^{s} h_{fk}\,\phi([y_j - u - u_k]/\sigma)\}$$

In the likelihood function, S is the number of possible genotypes at the QTL within the population (for example, it takes a value of 2 for a backcross population and 3 for an F_2 population); h_{fk} is the conditional probability of the individual having the k^{th} genotype at the QTL ($1 \le k \le S$) given its genotype f ($1 \le f \le R$) at the flanking marker loci, M_i and M_{i+1} (R = 4 or 9 for a backcross or an F_2 population respectively); and $\phi(\bullet)$ stands for the probability density function of a standard normal distribution. Calculation of the conditional probability h_{fk} depends on the genetic structure of the segregating population in question (i.e. F_2, backcross, etc.) and the location of the QTL within the flanking marker interval, and has been illustrated elsewhere (Lander & Botstein 1989; Luo & Kearsey 1992).

Differentiating the logarithm of the likelihood function with respect to each of the unknown parameters and setting the corresponding derivative to zero gives

$$\frac{\partial \ln L_q}{\partial x} = \sum_{j=1}^{n} \frac{1}{\sum_{l=1}^{S} h_{fl}\, \varphi([y_j - u - u_l]/\sigma)} \sum_{k=1}^{S} h_{fk} \frac{\partial}{\partial x} \varphi([y_j - u - u_k]/\sigma) = 0$$

Taking $x = u, u_k (k=1, 2,\ldots, S)$ and σ^2 in order, and solving the differential equations, yields the maximum likelihood estimates (MLE) of the model parameters given by

$$\hat{u} = \sum_{j=1}^{n} \sum_{k=1}^{n} w_{jk} (y_j - u_k)/n$$

$$\hat{u}_k = \sum_{j=1}^{n} w_{jk} (y_j - \hat{u}) / \sum_{j=1}^{n} w_{jk}$$

in which

$$w_{jk} = \frac{h_{fk}\, \varphi([y_j - u - u_k]/\sigma)}{\sum_{l=1}^{S} h_{fl}\, \varphi([y_j - u - u_l]/\sigma)}$$

represents the posterior probability of individual j having the k^{th} genotype at the QTL given its trait phenotype y_j and flanking marker genotype f. In terms of the basic model given earlier for the additive and dominance effects of a QTL these parameters translate to: u = m; u_1 = m – a; u_2 = m + d; u_3 = m + a.

The algorithm demonstrated above for calculating the MLEs consists of two steps. The first step calculates the posterior probability distribution of the missing information about the QTL genotype; it results in producing an expected value for the missing data. Thus, this step is termed E-step for 'expectation' step. The second step involves calculating the MLEs of the unknown parameters. It is achieved by making use of the posterior probabilities and is termed the M-step for 'maximization' step. These two steps of numerical calculation are iterated starting with initial guesses for the parameter values (e.g. u, u_k and σ^2) being, for example, the corresponding sample estimates, until the likelihood function converges at any given

prior criterion. A rigorous mathematical treatment of the algorithm can be found in Dempster *et al.* (1977).

The test statistic for the presence of the putative QTL at the given chromosomal position within the flanking interval is the likelihood ratio

$$LR = 2\ln\frac{L_q(\hat{u};\hat{u}_1,\hat{u}_2,..,\hat{u}_S;\hat{\sigma}^2)}{L_o(\hat{\hat{u}},\hat{\hat{\sigma}}^2)}$$

where $L_o(\bullet,\bullet)$ is the likelihood of the null hypothesis and is calculated at the sample mean and variance of the trait phenotypic records. The likelihood ratio is converted into the log-odds score (LOD) by $LOD = (\log_{10} e)LR/2 = 0.217LR$, which is asymptotically distributed as a chi-square distribution with $df = S - 1$ under the null hypothesis.

The test above can be carried out at any given chromosomal position that is bracketed by a pair of marker loci. Thus, the method represents a systematic procedure to search for QTL over the whole genome and reduces a multi-dimensional search problem for multiple QTL to one involving just a single dimension. Because the test is performed repeatedly at multiple locations, a practical problem arises of determining an appropriate threshold for declaring the presence of QTL at a genome scanning level. In other words, how to avoid false positives. Several methods have been proposed in the literature to determine the threshold. Lander and Botstein (1989) developed an asymptotic formulation, which was based on an Orenstein-Uhlenbeck diffusion process, for a genome-wide LOD score threshold. They suggested the value of the threshold should be between 2 and 3 to ensure a 5% overall false-positive error. Alternatively, Churchill and Doerge (1994) proposed a data-based numerical method, based on the theory of the permutation test, to determine the critical value of the significance test in the QTL analysis. A permutation test is a general numerical approach for calculating a significance threshold for a test statistic. The test statistic is calculated based on random permutations of the data, simulating random sampling of the data under the null hypothesis that there is no QTL. This procedure is repeated a large number of times, resulting in a series of values of the test statistic under the null hypothesis. The $1 - \alpha$ percentile of these observed values of the test statistic gives the critical value (i.e. the threshold) at significance level α. The theoretical basis behind this approach was explained in detail in Lehmann (1986). The permutation test models the null hypothesis by essentially de-coupling the trait and marker data and deriving an empirical distribution of the test statistic.

The above analysis provides point estimates of QTL map locations, genetic effects and residual variances of the QTL. In practice, it is very important to know the sampling variances associated with these estimates, that is, how accurate they are. In order to estimate the confidence interval of a QTL map location, Lander and Botstein (1989) proposed the use of a one LOD support interval based on the asymptotic chi-square distribution of the likelihood ratio test statistic under the null hypothesis. Mangin *et al.* (1994) pointed out that this method is appropriate only when the QTL effect is large. However, when the QTL effect is small, the test statistic

may not follow a chi-square distribution and, as result, the one LOD support interval underestimates the confidence interval. They developed a novel statistic with asymptotic properties that do not depend on the QTL effect. In this, they were the first to provide an unbiased and feasible method for estimating the confidence interval of a QTL map location. In addition, Visscher *et al.* (1996) suggested calculating the sampling variances of the parameter estimates by making use of a bootstrap sampling method. Bootstrapping simulates the estimation of the QTL parameters from many repeats of the experiment. It achieves this by repeatedly sampling from the existing experimental trait and marker data with replacement, every sample mimicking a repeat of the experiment (Davison & Hinkley 1998). The statistical analysis is similarly performed on each replicate data set, yielding repeated estimates of the parameters. Variances of these repeated estimates provide the sampling variances of the estimates. More recently, Kao and Zeng (1997) formulated the observed information matrix of the MLEs of the model parameters in an interval mapping analysis. In theory, the inverse of the matrix provides the corresponding estimates of the sampling variances. However, Luo *et al.* (2000) showed that this approach might fail to produce stable estimates of the sampling variance when convergence of the 'expectation and maximization' (EM) algorithm was slow, and an approximate but robust method for calculating the stable estimates was proposed.

The interval mapping theory is built as the kernel of the statistical framework for QTL mapping. Since its publication, many modified versions based on the basic principle have been proposed for improving various aspects of this method. A simple regression model was suggested in Haley and Knott (1992). Instead of treating the quantitative trait as a mixture model with missing information, the regression analysis models individual phenotypic record y_j by

$$y_j = u + \xi_{jk} u_k + e_j$$

in which ξ_{jk} is the conditional probability of the individual having the k^{th} QTL genotype given its genotype at the flanking marker loci and the test position of the putative QTL. Statistical analysis of the model is straightforward, and it provides the flexibility to fit different fixed effects in the model such as site, sex and maternal effects. Numerical analyses based on simulation studies have shown that this procedure gives a very good approximation to the likelihood-based analysis even though it was pointed out by Xu (1997) that the regression approach tends to overestimate the residual variance.

One practical problem of interval mapping with a pair of flanking markers is that the test statistic at any test position will be affected by other QTL linked or unlinked to the test position. To overcome this problem, Zeng (1994) and Jansen and Stam (1994) suggested the use of other markers in addition to the flanking markers as a background control in the interval mapping analysis. The composite interval mapping for such analysis combines interval mapping with multiple regression and fits the following linear model for quantitative trait value of the j^{th} individual as

$$y_j = u + u_k + \sum_{i=1}^{m} b_i x_{ji} + e_j$$

where x_{ji} is the type of the i^{th} marker in the individual and b_i is the partial regression coefficient of phenotype y on the marker i conditional on all other markers. The consequences of incorporating additional m markers as cofactors into the model depend on the relationship between these markers and the flanking markers. The role played by the cofactor markers is similar to that of covariates in multiple regression analysis. When the additional markers are linked to the flanking markers, the effect of those QTL located outside the boundaries defined by these markers will be effectively controlled. This leads to a higher mapping precision, but at the same time the statistical power for detecting the QTL may be reduced under the conditional test (Zeng 1993). However, use of unlinked markers as cofactors in the model may be effective in reducing residual variation of the model, and in turn leads to increase in the test power.

A common feature of the interval mapping and the composite interval mapping protocols is their one-at-a-time strategy, whereby one evaluates the association of a single QTL with a marker interval while ignoring the interaction of the tested QTL and other QTL segregating in the mapping population. Because epistatic effects between different QTL are found to be a common phenomenon for most quantitative complex traits (Mackay 2001a), these one-at-a-time approaches are limited in detecting such effects. To meet this requirement, Kao *et al.* (1999) proposed a multiple QTL interval mapping approach, which considers all possible parameters defining the genetic architecture of polygenic inheritance in a likelihood-based statistical analysis. These include the number, effects and epistasis of QTL, genetic variance and covariances explained by QTL effects. Implementation of the multiple interval mapping is computationally much more demanding than other interval mapping approaches but its dynamic search for all significant genetic components makes the model fitting more close to the real genetic architecture of quantitative traits.

It must be pointed out that the interval mapping and its modified versions share the common problem that estimates of QTL map locations and effects are highly model-dependent. The use of a misleading model may result in severely biased prediction of the genetic parameters. To avoid this malpractice, QTL mapping analysis should not be built solely on a simple additive/dominance model but integrate all aspects of genetic analysis of quantitative traits (Mackay 2001b).

Efficiency of QTL mapping analysis may be influenced by many factors, predominant among which are population size and trait heritability. The density of markers scored in a mapping experiment defines the scale of map information, but there exists an upper limit to improvement in the efficiency of QTL mapping through increasing marker density. The use of an extremely dense marker map will be of little use if the mapping population fails to provide sufficient recombination between the markers and QTL (Hyne *et al.* 1995). Moreover, the segregating populations derived from different mating designs can be characterized with different genetic structures and thus exhibit varying utilities in QTL mapping. For example, use of backcross populations is more powerful than use of F_2 populations for detecting QTL effects,

but the latter are preferred for achieving better estimates of the effects (Darvasi 1997).

1.4.3 QTL mapping in pedigree populations

In many out-breeding species the establishment of inbred lines is not practical. Mapping QTL in these species cannot be performed by use of simple segregating population as discussed above. Populations of these species (e.g. most trees) exist in a pedigree structure. In sharp contrast to the segregating populations, mapping genes segregating in the pedigree population is much more problematic. First, the size of the nucleus family in most pedigree populations is substantially smaller than commonly used segregating populations in QTL analysis. The whole analysis involves a large number of independent pedigree families to ensure an adequate statistical power for detecting linkage between markers and QTL. Second, information about the linkage phase of genes at the marker loci and QTL is no longer directly extractable from these pedigrees. To achieve this requires the development of complicated statistical tools for modelling the inheritance of genes within a multiple generation pedigree and sophisticated computational algorithms to assess the likelihood of all possible configurations of linkage phases at a finite number of loci. Guo and Thompson (1992) developed a Gibbs sampling-based approach to combine conventional segregation analysis at individual loci with linkage analysis. Gibbs sampling is a numerical approach to calculate marginal distributions from joint distribution (Casella & George 1992). In the linkage analysis setting, this technique was used to calculate the likelihood function by integrating over the polygenic additive effects. This provides a test for, and estimates of, the linkage between a single marker locus and a locus underlying quantitative genetic variation in a large complicated pedigree. Third, the fact that different families may bear different alleles at QTL poses another severe problem of genetic heterogeneity in the linkage analysis. All of these make QTL mapping in pedigree populations a challenging topic in both theoretical and experimental studies.

There are several approaches of QTL analysis that make use of multiple pedigrees with a simple consanguineous relationship among members. The basic idea of the interval mapping of QTL was extended by Fulker and Cardon (1994) to model a series of independently collected sib-pairs. Instead of working with the probability of QTL genotype conditional on flanking marker genotype, they modelled the proportion of genes IBD (identical by descendent) shared by a sib-pair at a putative QTL (Π_q) in terms of the IBD genes shared at two flanking marker loci (Π_1 and Π_2) as

$$\Pi_q = \alpha + \beta_1 \Pi_1 + \beta_2 \Pi_2$$

in which

$$\beta_1 = [(1-2c_1)^2 - (1-2c_2)^2(1-2c)^2] / [1-(1-2c)^4]$$

$$\beta_2 = [(1-2c_1)^2 - (1-2c_2)^2(1-2c)^2]/[1-(1-2c)^4]$$

$$\alpha = (1-\beta_1-\beta_2)/2$$

where c_1, c_2 and c are respectively recombination frequencies between the left flanking marker and the QTL, between the QTL and the right flanking marker, and between the flanking marker loci. Regression of the difference in trait phenotype record between the sib-pair on the IBD proportion at the putative QTL (Π_q) creates a test statistic for detecting the presence of the QTL. The analysis can be performed at any position within the flanking interval, and thus the method essentially provides a systematic search for QTL. Simulation studies suggested that thousands of sib-pairs are needed in order to obtain adequate power to detect the QTL and meaningful estimates of the QTL parameters. However, the efficiency of the design should be improved substantially if the sib-pairs with extreme phenotype are selectively used in the QTL mapping analysis (Risch & Zhang 1995).

1.4.4 QTL analysis in natural populations

The precision with which a single QTL can be localized relative to a marker locus is directly proportional to the number of informative meioses provided by a mapping population. A large number of such informative meioses indicate a large number of recombinations between the marker and trait locus. The linkage analyses demonstrated in the previous two sections are typically restricted in the number of such useful meioses, and thus have yielded poor mapping resolution. It was observed that the 95% confidence interval for QTL map locations inferred from many plant experiments were often in a range of 20 to 30 cM and seldom less than 5 cM (Kearsey & Farquhar 1998). This is too coarse for utility of the mapping information in marker-assisted selection for genetic improvement of quantitative traits or in targeting candidate genes affecting the traits. Use of historically accumulated recombinations in the populations from a well-designed breeding scheme (Xiong & Guo 1997) or in natural populations (Lander & Schork 1994) has been shown to be an effective way to improve resolution of QTL mapping. Methodologically, use of natural populations for mapping genes underlying quantitative traits requires modelling linkage disequilibrium between genes segregating at marker loci and trait loci.

Linkage disequilibrium is a central concept of population genetics and is defined as the non-random association of alleles at different loci in a given population. The degree of the non-random association is referred to as the coefficient of linkage disequilibrium (Crow & Kimura 1970). The basic idea behind the linkage disequilibrium analysis for gene mapping has been demonstrated in Terwilliger (1995) and Kaplan and Weir (1997). Suppose that a mutation at a gene affecting a character occurred many years ago and – possibly through a founder effect – was propagated in the population. Thus, it is possible that the marker alleles on the original mutant haplotype may still be in linkage disequilibrium with the mutant allele in the

current population. If so, the level of the disequilibrium is likely to be proportional to the number of generations since the mutation occurred and to the genetic distance (i.e. the amount of recombination) between the marker and the mutant locus. If a significant disequilibrium is detected in the current population, a larger generation number may imply a closer linkage between the marker and mutant alleles.

Linkage disequilibrium has a long and important history in population genetics theory, but linkage disequilibrium analysis for quantitative trait loci is a matter of recent interest in genomics. The difficulties encountered in modelling linkage disequilibrium involved with QTL are mainly caused by the unavailability of genotypic data on the trait. Luo (1998) presented a theoretical model for linkage disequilibrium between a polymorphic marker locus and a locus affecting quantitative genetic variation in a natural population. The model considers two bi-allelic loci: one affects a quantitative trait (QTL), whereas the other is a co-dominant marker that is devoid of effect on the trait. The two alleles are denoted by *M* and *m* at the marker locus and *A* and *a* at the QTL. Three genotypes at the QTL, say, *AA*, *Aa* and *aa*, are assumed to have the genotypic values $G_{AA} = u + a - d/2$, $G_{Aa} = u + d/2$ and $G_{aa} = u - a - d/2$ respectively, where *a* and *d* are the additive and dominance effects of QTL. The association between the two loci in quantitative genetic effect is quantified by linkage disequilibrium with *D*, the coefficient of the disequilibrium, defined as $D = f_{MA} - pq$, where f_{MA} is the frequency of the *MA* haplotype and *p* and *q* denote frequencies of alleles *M* and *A* respectively. With the assumption of random mating, the joint distribution of genotypes at the marker locus and QTL in a natural population can be expressed as a function of *D*, *p* and *q*, and is shown in Table 1.1.

Based on the population genetics model given in Table 1.1, the phenotypic value of individual *k* in the population, which has marker genotype *i* and the underlying QTL genotype *j*, can be modelled as

$$y_{ijk} = u + \beta_i + \omega_{ij} + \varepsilon_{ijk}$$

Table 1.1 Joint distribution of marker and QTL genotype in a random mating population: where *D* is the coefficient of linkage disequilibrium between the marker and the QTL; *p* and *q* are the frequencies of marker allele *M* and QTL allele *A*; *a* and *d* are the additive and dominance effects of QTL.

	QTL genotype		
Marker genotype	AA	Aa	aa
MM	$(D+pq)^2$	$2(D+pq)[p(1-q)-D]$	$[D-p(1-q)]^2$
Mm	$2(D+pq)[(1-p)q-D]$	$2[2D^2+(1-2p)(1-2q)D$ $+2pq(1-p)(1-q)]$	$2[D+(1-p)(1-q)]$ $[p(1-p)-D]$
mm	$[D-(1-p)q]^2$	$2[(1-p)q-D][D+(1-p)(1-q)]$	$[D+(1-p)(1-q)]^2$
Genotypic value	$u+a$	$u+d$	$u-a$

where i = 1,2,3 refers to the marker genotypes *MM*, *Mm* and *mm*, and j = 1,2,3 to the QTL genotype *AA*, *Aa* and *aa* respectively. u is the population mean, β_i is the QTL effect associated with the marker genotype i, ω_{ij} is the effect of the j^{th} QTL genotype within the i^{th} marker genotype, and ε_{ijk} is the residual random effect whose distribution is assumed to be normal with mean zero and variance σ^2.

If a random sample of n individuals is collected from the population, the data of trait phenotype and marker genotype scored on these individuals may be analysed in three different ways.

1.4.4.1 Analysis of variance under an unbalanced nested design

Expressed in terms of parameters in the above linear model, it has been shown by Luo (1998) that the expected between marker genotype effects are given by

$$\beta_1 = \frac{2D[pa-(D-p+2pq)d]}{p^2}$$

$$\beta_2 = \frac{D\{(1-2p)a+[2D+(1-2p)(1-2q)]d\}}{p(1-p)}$$

$$\beta_3 = -\frac{2D\{(1-p)a+[D+(1-p)(1-2q)]d\}}{(1-p)^2}$$

These show that significant variation between the marker genotypes on the QTL effect is an indicator of the presence of linkage disequilibrium between the marker and QTL. To perform a theoretical analysis of the model and to take into account that the number of individuals with marker-QTL genotype ij, say n_{ij}, is a random number from a multinomial distribution described in Table 1.1, Luo (1998) worked out that the expected mean square between the marker genotypes is given as:

$$\begin{aligned} EMS_\beta = {} & \frac{1}{(G-1)}\{\sum_{i=1}^{G} nf_i\beta_i^2 - [\sum_{i=1}^{G} f_i(1+(n-1)f_i)\beta_i^2 + 2(n-1)\sum_{i<j\le 3} f_i f_j \beta_i \beta_j] \\ & + \sum_{i=1}^{G}\frac{1}{f_i}[\sum_{i=1}^{G} f_{ij}(1+(n-1)f_{ij})\omega_{ij}^2 + 2(n-1)\sum_{i<k\le 3} f_{ij}f_{ik}\omega_{ij}\omega_{ik}] \\ & - [\sum_{i=1}^{G}\sum_{j=1}^{3} f_{ij}(1+(n-1)f_{ij})\omega_{ij}^2 + 2(n-1)[\sum_{i<j\le 3}\sum_{k<l\le 3} f_{ij}f_{kl}\omega_{ij}\omega_{kl}] + \sigma^2 \end{aligned}$$

and the expected mean square within the marker genotypes is:

$$\begin{aligned} EMS_\omega = {} & \frac{1}{n-G}\{\sum_{i=1}^{G}\sum_{j=1}^{3} nf_{ij}\omega_{ij}^2 - \sum_{i=1}^{G}\frac{1}{f_i}[\sum_{i=1}^{3} f_{ij}(1+(n-1)f_{ij})\omega_{ij}^2 \\ & + 2(n-1)\sum_{i<k\le 3} f_{ij}f_{ik}\omega_{ij}\omega_{ik}]\} + \sigma_e^2 \end{aligned}$$

where G is the number of marker genotypes observed in the sample, f_{ij} is the frequency of the j^{th} QTL genotype and the i^{th} marker genotype and is given in Table 1.1. f_i, the frequency of the i^{th} marker genotype.

1.4.4.2 Regression analysis

If the trait phenotype of individual k is regressed against x_k, the number of marker alleles M it carries, the regression model has a form of

$$y_k = u + bx_k + e_k$$

with the regression coefficient given as

$$b = \frac{D[a + (1-2q)d]}{p(1-p)}$$

A test for significance of the regression coefficient is essentially a test for significance of the coefficient of linkage disequilibrium. The sampling variance of the regression coefficient has a complicated form if the nature of discrete variate, x_k, is appropriately accounted for in formulating the variance. This was also given in Luo (1998).

The statistical properties of the analyses based on the ANOVA and the regression models were investigated in detail by Luo (1998) and Luo and Suhai (1999).

1.4.4.3 Likelihood analysis

Under the theoretical model given by Table 1.1, the likelihood of the population genetic parameters $\Omega = (p, q, D, u, a, d, \sigma^2)$, given the observed trait data (Y) and the marker genotypic data (X), can be written as

$$L(\Omega, Y, X) = \prod_{j=1}^{n} \Pr ob\{y_j | x_j, \Omega\} \Pr ob\{x_j | \Omega\}$$

where $Prob\{y_j|x_j,\Omega\}$ is the conditional probability of y_j given the marker genotype and the parameters Ω, and $Prob\{x_j|\Omega\}$, the conditional probability of the marker genotype given the parameters. If the data are sorted into classes according to the marker genotype and QTL genotype, say y_{ij}, and assume y_{ij} follows a normal distribution with mean u_k and variance σ^2, the log-likelihood can be simplified into a form given by

$$\log[L(\Omega, Y, X)] = \sum_{i=1}^{3} \sum_{j=1}^{n_i} \log[\sum_{k=1}^{3} f_{ik} e_{ijk}] - \frac{n}{2} \log(2\pi\sigma^2)$$

where $e_{ijk} = exp[-(y_{ij}-u_k)^2/2\sigma^2]$, $u_k = u+(2-k)a+(-1)^k d/2$ and $n_1+n_2+n_3=n$ with n_i being the number of individuals with the i^{th} marker genotype.

Luo *et al.* (2000) developed the EM algorithm to calculate the maximum likelihood estimates of the unknown parameters. Based on the likelihood analysis,

the test for the null hypothesis, H_o: $D=0$ vs. the alternative hypothesis, H_1: $D \neq 0$ may be performed through a likelihood ratio,

$$LR = 2\{\log[L(\hat{\Omega}, Y, X)] - \log[L(\hat{\Omega}, Y, X)]|_{\hat{D}=0}\}$$

The first part of the test statistic is the maximum log-likelihood under the full model (the alternative hypothesis) and the second is that under the reduced model with D restrained to zero (the null hypothesis). Details for the likelihood analysis are presented in Luo *et al.* (2000).

The above analyses are appropriate for those traits exhibiting continuous phenotypic distribution. Luo and Wu (2001) extended the theory and methods to predict linkage disequilibrium between a polymorphic marker locus and a locus affecting a complex trait whose phenotypic distribution is dichotomous.

The major difficulty in linkage disequilibrium-based mapping is to quantify the relationship between mapping distance and linkage disequilibrium measure. Because recombinant events are not observed, the recombination fraction between the marker and trait locus must be estimated on the basis of a population genetics model. Several methods have been suggested to address this problem. One of these searches for the reparameterization by which the disequilibrium measure can be directly related to the recombination fraction. For instance, Devlin and Risch (1996) found that, in the present notations, the measure of the disequilibrium

$$\delta = \frac{D}{q[(1-p)(1-q)+D]}$$

is related to the recombination fraction θ as $\delta = (1-\theta)^T$, where T represents the generation number since the creation of the initial disequilibrium. T may be estimated either from an epidemiological survey (i.e. in Hastbacka *et al.* 1992) or directly from the sampled data (i.e. Kaplan *et al.* 1995; Thompson & Neel 1996). With this estimate – together with those from the present analyses – the recombination fraction is calculated from

$$r = 1 - \exp\left\{\frac{1}{\hat{T}}\log\left[\frac{\hat{D}}{min(\hat{p}(1-\hat{q}),(1-\hat{p})\hat{q})}\right]\right\}$$

Figure 1.1 illustrates the recombination fraction between two loci predicted from the level of linkage disequilibrium, which is maintained over various generations since the establishment of initial disequilibrium in a random mating population. It demonstrates that the evolutionary processes play an effective role in breaking down linkage between two loci, and that linkage disequilibrium may be sustained for many generations only between genes that are closely linked. This provides an opportunity to locate genes affecting a complex trait at a resolution that is rarely achievable using segregating populations.

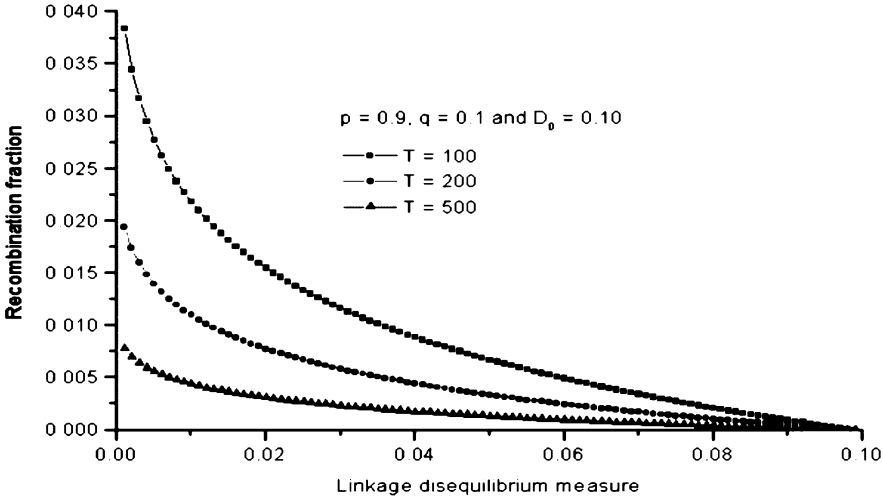


Figure 1.1 Relationship between map distance and linkage disequilibrium between two loci in a random mating population.

A linkage disequilibrium-based strategy has been successfully implemented in mapping candidate genes for several inherited diseases in human populations, and some of them have identified the disease genes (e.g. Hastbacka *et al.* 1994). The approach has started to appear in the literature for evaluating the association of polymorphisms within well-characterized candidate genes for quantitative traits in agricultural crops (Thornsberry *et al.* 2001). The power of this approach is fully realized from using extremely dense maps of polymorphic markers (single nucleotide polymorphisms, SNPs) because they allow genome-wide screening for those polymorphisms that are in significant disequilibrium with the trait. The analyses presented here are restrained to a biallelic model at both marker and trait loci. Multiple alleles at the marker loci may need to be taken into account for two reasons. First, multiple alleles are common in natural populations for molecular DNA markers like microsatellites. Second, the multiple-allele model enables the problem of linkage disequilibria between multiple marker loci and a trait locus to be reduced to the problem of two locus analysis by treating haplotypes over a group of markers as 'multiple alleles'. Thus, the model and analysis represented here may be integrated into a multiple marker framework when they are extended to take multiple alleles into account.

The above sections discussed the basic principles and methods of the major streams of QTL mapping. The relevant statistical algorithms have been developed into a wide range of computer packages for QTL mapping, many of which have been made fairly user-friendly for experimental biologists. An overview of these packages together with a list of their web sites is available in Manly and Olson (1999).

1.5 The application of QTL theory

In order to study QTL for traits relevant to plant breeding, the traits generally need to be measured under standard agricultural conditions. This normally implies using plots to simulate the field situation. Raising F_2s or backcrosses would not be appropriate because they would have to be raised as spaced plants in order to identify and score them both for markers and traits. Furthermore, unless they could be clonally propagated, the same plants could not be raised and scored in different years or locations. It would, therefore, be necessary to use different genotypes each time and so would require repeated marker genotyping, which is inefficient and error prone. The solution to these problems is to use genotypes that are genetically more uniform and examples of these are presented in the following sections.

1.5.1 Advanced segregating populations

Typical examples would be recombinant inbred lines (RILs), produced by single seed descent (SSD) from an F_2, early generations of selfing such as F_3 or F_4 families, and doubled haploid lines (DHLs) derived by microspore culture from an F_1. Apart from the F_3 and F_4, these are all true breeding generations, so-called 'immortal' populations, and so seed can be multiplied up to any degree for large-scale replicated plot trials. They also need to be genotyped once only. RILs have an extra advantage because they allow additional rounds of recombination that break up tight linkages. The same principles as described above can be used to analyse data from such collections of lines, to locate QTL and estimate their effects, the formulae simply have to be adjusted to allow for the different levels of recombination.

Such 'immortal' populations have been used to locate QTL in most crop species (for reviews see Tanksley 1993; Kearsey & Farquhar 1998) and provide sufficiently accurate location for marker-aided selection. Sometimes these locations give some clue to the mode of action of the genes involved. For example, Bohuon *et al.* (1998) located six QTL affecting flowering time in a *Brassica oleracea* population, three of which mapped to syntenous regions of chromosomes 2, 3 and 9 (*B. oleracea* is an ancient polyploid). The QTL locations in these three syntenous regions also overlapped a region syntenous to part of *Arabidopsis* chromosome 5 containing the zinc-finger protein gene, *CONSTANS*. Similar regions syntenous with *CONSTANS* have also been implicated in flowering time QTL in *B. nigra* and *B. juncea* (Lagercrantz *et al.* 1996; Axelsson *et al.* 2001; see also Chapter 3). The very close linkage of the flowering QTLs to *CONSTANS,* rather than to an alternative candidate, *FLC*, has been demonstrated by Axelsson *et al.* (2001). Yano *et al.* (2000) have identified a QTL for heading date in rice (*HD1*) that has also been shown (by map-based gene cloning) to correspond to *CONSTANS*. These types of evidence, together with others discussed later (e.g. Frary *et al.* 2000), strongly suggest that much QTL allelic variation may be regulatory. It is believed that the lines to *Arabidopsis* and *Brassica* diverged 12 to 19 million years ago (Cavell *et al.* 1998), and those to rice ~200 million years ago (Wolfe *et al.* 1989); hence, the genes have been strongly conserved.

It is well established that DHLs can exhibit distorted segregation of markers along various segments of a chromosome. This is due to allelic differences that affect success in microspore culture and/or plant regeneration from callus. These differences cause selection for responsive alleles. This selection, in turn, affects the frequencies of alleles at linked marker loci, causing the distorted ratios to extend over tens of cM along a chromosome with a peak near the offending gene. Similar selective effects can also occur in RILs, although the reasons are less clear. Such distortions will also affect the overall trait mean of the population of DHLs or RILs if QTL are within these regions. Biases of this sort will resemble the effects of epistasis, such as when the overall mean of the DH lines transgresses the parents.

Such selective distortions have no serious effect on mapping, unlike distortions due to errors in genotyping which result in aberrant ratios at single, isolated markers. They appear as rare double recombinants and so create an upward bias in map length. They are caused by mis-scoring of ambiguous gel images, typographical errors or by temporary methylation affecting access to nuclear sites. The latter effect behaves like a single point mutation but is not necessarily inherited. It is essential to check for and eliminate such errors before any QTL analysis is considered. All ambiguous data must be discarded.

1.5.2 Part chromosome substitution lines (backcross introgression lines) and near-isogenic lines (NILs)

Unless population sizes are very large, the confidence intervals (CIs) associated with QTL location can also be large. The lower the heritability of the trait and the smaller the QTL effect, the larger the CI. For typical population sizes of RILs used by breeders of 100 to 200 lines, CIs for detectable QTL are seldom less than 5 cM and often 20 cM or more (Kearsey & Farquhar 1998). Having additional markers above one every 10 cM has no significant advantage in reducing the CI. To make serious reductions in CI requires the population size to increase exponentially, but this is impractical on a field trial basis. There are also the potentially serious problems to plant breeding of false-positive QTL, and it is always advisable to confirm locations using replicate testing populations, as is well argued by Melchinger and collaborators (Melchinger *et al.* 1998, 2000; Utz *et al.* 2000). Nonetheless, such population-based studies do provide a useful initial approach to identifying the approximate location of key QTL, and such accuracy may be adequate for MAS.

However, other approaches are needed to provide the more precise location needed for candidate gene identification and cloning. In deciding on strategies for more accurate QTL location, it is important to draw a clear distinction between what is possible and what is practical in a plant-breeding context. Ultimately this has to be an economic decision and is beyond the scope of this chapter; so we will concentrate on what is possible. The principle behind all approaches is to use genetic material that places defined, narrow limits on the possible region of chromosome containing the QTL. Ideally, this involves producing two genotypes that are genetically identical except for a short chromosomal region and demonstrating that they differ for the

trait in question. Obviously, the smaller the region, the more precisely the QTL can be located. Such approaches have the advantage of concentrating trait scoring to just two (or a very few) lines allowing highly replicated and hence very powerful trials. Three different approaches have been used – part chromosome substitution lines (CSLs); near-isogenic lines (NILs); and stepped aligned inbred recombinant strains (STAIRS) – and we present their construction, analysis and relative merits below. All are normally derived from two defined, inbred parental lines chosen because they differ for traits of interest.

1.5.3 Part CSLs

These consist of a standard inbred line or variety into which short regions of chromosome from a donor line have been introgressed by MAS. Typically, construction involves backcrossing the F_1 (between the donor and recipient variety) to the recipient line for one or more generations. Molecular markers are used to identify individuals that are homozygous for recipient marker alleles for most of the genome but heterozygous for just one or two short regions. Such individuals are selfed, and those progeny that are homozygous for the donor alleles in these regions selected and selfed again to multiply seed of that substitution (Howell *et al.* 1996; Kearsey & Pooni 1996; Ramsay *et al.* 1996; Tanksley & Nelson 1996; Monforte & Tanksley 2000a; Monforte *et al.* 2000b). In practice, a large number of different substitutions are generated, varying in length and location. The logistics of the MAS often requires that relatively few markers are used with large inter-marker distances. So, subsequently defining the precise cross-over positions between donor and recipient chromosomes can be difficult and time-consuming and may only pursued with those substitutions that show useful QTL activity.

Lines with these substitutions are raised in trials and those that differ significantly from the recipient variety are identified as potentially containing QTL in the introgressed region. It is wise to repeat the trial with those lines identified as carrying QTL in order to avoid false positives. Comparisons among lines with overlapping substitutions can often enable the QTL location to be further restricted to part of a given substitution (Rae *et al.* 1999; Monforte & Tanksley 2000a; Burns *et al.* 2002). The analytical principles are analogous to deletion mapping used to locate genes in many organisms including mice and man. The substituted region can be reduced still further to provide a narrower confidence interval for the QTL, or the QTL location can be identified by backcrossing to the recipient line and scoring progeny that have recombined within the substituted region using an approach described by Mather and Jinks (1982, pp. 16–17) and Kearsey and Pooni (1996, pp. 168–171). However, this requires scoring individual plants and so may not always be appropriate to crop species.

1.5.4 NILs

The aim of NILs is the same as for CSLs, but they are developed differently. If an F_2 is inbred by SSD, each SSD line will become increasing homozygous at all loci,

the proportion of heterozygous loci halving at each generation. By F_6, for example, a given individual may be completely homozygous at all loci except for a tract of chromosome that remains heterozygous. If this individual is selfed, the previously heterozygous region will segregate like a single unit, apart from rare recombination, resulting in three genotypes – the two alternative homozygotes and the heterozygote. The progeny can be scored for the trait and the markers in the critical region; any differences between the three genotypes in the trait can be thus ascribed to QTL in that region. Alternatively, the two opposing homozygotes can be selfed to fix the alternative 'alleles' of the critical region for future study. Thus, pairs of lines are produced which are identical apart from the particular critical region. The lines are identified using molecular markers and, as with CSLs, the ends of the critical region can be determined by further genotyping. NILs can also be produced by backcrossing, but extra work is required to achieve the crosses as opposed to selfs.

NILs have two advantages and two disadvantages over CSLs. On the plus side, they are easier to construct and it is easier to produce quite small substitutions because there have been more rounds of recombination. On the minus side, one does have to rely on chance to produce the NILs containing a region containing a QTL of interest, but attempts are being made to construct a genome-wide library of NILs in some key species. The second disadvantage is that each pair of NILs has a different genetic background to all other NILs, though this can be overcome by generating NILs by backcrossing rather than selfing.

NILs have become a regular feature of QTL characterization leading to map-based gene cloning. For example, Monforte *et al.* (2001b) have used introgressions from three wild tomato species to locate QTL affecting fruit quality and yield-related traits on chromosome 4. They were able to show the relative unimportance of genotype by environment or epistatic effects involving these genes, so making the NILs a useful resource for gene introgression. They were also able to demonstrate that some QTL involved several closely linked genes which would appear as single QTL by population-based methods. Van Berloo *et al.* (2001) were able accurately to locate a gene responsible for leaf rust (*Puccinia hordei*) resistance in barley and make the segment available for marker-assisted selection. Lin *et al.* (2000) were able do locate three QTL for heading date using NILs in rice by introgressing donor chromosomal segments from the *indica* variety Kasalath into Nipponbare (*japonica*). They were also able to explore the epistatic and environmental interactions between these QTL-NILs and concluded that the QTL had a regulatory role. In *Arabidopsis*, El-Assal *et al.* (2001) were able to identify a novel cryptochrome allele of *CRY2* which is responsible for QTL variation for flowering time.

CSLs and NILs are a useful resource to provide a location for QTL for traits that are difficult to record on individual plants such as transformation ability and disease susceptibility. For example, Cogan *et al.* (2002) used a *Brassica oleracea* DHL population to map QTL controlling the *Agrobacterium rhizogenes*-mediated adventitious and transgenic root production. The parental genotypes used to produce these DHLs had previously been used to produce sets of substitution lines (Ramsay *et*

al. 1996), and these were employed to confirm the map positions and more accurately locate the QTL identified using the DHLs (H.J. Newbury, pers. commun.).

1.5.5 STAIRS

These resources are designed to achieve what NILs provide, but in a more targeted way, and also to provide the opportunity for much smaller CIs. However, they rely on the initial availability of whole chromosome substitutions from a donor into a recipient and so far these are available only in a few model organisms such as *Arabidopsis*, wheat, *Drosophila* and mice. For example, there would be five such substitutions of all possible single donor chromosomes in *Arabidopsis* into a recipient, and these are produced by marker-aided selection and backcrossing.

If the F_1 between one such chromosome substitution strain (CSS) and the recipient line is crossed again to the recipient, a population of first-backcross progeny can be obtained. These will contain a mixture of chromosomes from the F_1 parent with zero, one, two, three or more cross-overs; of these, 50% will contain just a single cross-over. They can be identified using a small number of markers covering the length of the target chromosome, and each recombinant chromosome can be fixed by selfing and marker selection to produce single recombinant lines (SRLs). With 1000 such SRLs, there would be 10 different single recombinants for every 1 cM on average. If these are aligned according to the position of the cross-over, they provide a series of steps along the chromosome, any two steps differing by just 0.1 cM. These stepped aligned inbred recombinant strains (STAIRS) can be used for gene (QTL) location, allowing one to close in on a gene of interest as follows.

Initially, one would score just a limited number of lines for the trait using wide steps up the STAIRS to identify approximate regions containing QTL. One would then examine narrower-stepped STAIRS within this region to provide a narrower focus, finishing with very narrow (~0.1 to 0.5 cM) STAIRS to provide fine resolution (Kearsey 2002; Koumproglou *et al.* 2002).

1.5.6 Cloning QTL

As we have seen, CSLs and NILs enable QTL to be mapped to relatively small regions of a chromosome, delineated by molecular markers. In those species for which relatively complete sequence information is available (e.g. *Arabidopsis* and rice), such markers can point directly to physical regions of the chromosome allowing in-silico bioinformatics procedures to identify possible candidate loci. In non-sequenced species, these markers can be used to identify clones in either bacterial or yeast artificial chromosome cDNA libraries (BACs or YACs). These clones can be organized into contigs and sequenced. At this point, conventional bioinformatics tools can be used to search for open reading frames and to identify potential candidates from homology with genes of known function in other species.

Frary *et al.* (2000) were able to identify a cosmid-containing DNA taken from a wild tomato species in the region of a QTL with a very large negative effect on fruit

size (fw2.2). This cosmid was transformed into cultivars with large fruits, and it reduced the size by the amount expected from the QTL effect. The gene appears to have a regulatory rather than structural function and controls the carpel cell number during early floral development. Such a regulatory function was also inferred for heading time QTL in *Brassica* (Bohuon *et al.* 1998) and rice (Lin *et al.* 2000) as described earlier. Kamolsukyunyong *et al.* (2001) were able to locate a QTL involved with submergence tolerance in rice to a 6-cM interval of chromosome 9. Using YAC and BAC clones, they were able to narrow this down to a region of very high recombination frequency. El-Assal *et al.* (2001) cloned a candidate gene for flowering time in *Arabidopsis*, and transformants showed the phenotype expected.

1.6 Conclusion

QTL analysis with molecular markers has enormously accelerated the development of quantitative genetics during the past 15 years. It is now possible to obtain approximate locations for agronomic QTL, estimate their effects, and bracket them with markers to facilitate marker-assisted introgression and breeding. Finer QTL mapping with NILs and other introgressed segments can now lead to accurate candidate gene location, cloning and transformation in key crop and model species. These techniques offer considerable promise for future plant breeders.

References

Axelsson, T., Shavorskaya, O. & Lagercrantz, U. (2001) Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. *Genome*, **44**, 856–864.

Bohuon, E.J.R., Ramsay, L.D., Craft, J.A., *et al.* (1998) The association of flowering time QTL with duplicated regions and candidate loci in *Brassica oleracea*. *Genetics*, **150**, 393–401.

Breese, E.L. & Mather, K (1957) The organisation of polygenic activity within a chromosome in Drosophila: I Hair characters. *Heredity*, **11**, 373–395.

Breese, E.L. & Mather, K. (1960) The organisation of polygenic activity within a chromosome in Drosophila: II Viability. *Heredity*, **14**, 375–400.

Burns, M.J., Barnes, S.R., Bowman, J.G., Clarke, M.H.E., Werner, C.P. & Kearsey, M.J. (2003) QTL analysis of an intervarietal set of substitution lines in *Brassica napus*: I Seed oil content and fatty acid composition. *Heredity*, **90**, 39–48.

Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. *American Statistical Association*, **46**, 167–174.

Cavell, A.C., Lydiate, D.J., Parkin, I.A.P., Dean, C. & Trick, M. (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome*, **41**, 62–69.

Chicurel, M. (2001) Faster, better cheaper genotyping. *Nature*, **412**, 580–582.

Churchill, G.A. & Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971.

Cogan, N.O.I., Lynn, J.R., King, G.J., Kearsey, M.J., Newbury, H.J. & Puddephat, I. (2002) Quantitative trait analysis associated with *Agrobacterium rhizogenes*-mediated adventitious and

transgenic root production in *Brassica oleracea*. *Theoretical and Applied Genetics*, **105**, 568–576.
Crow, J.F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory.* Harper & Row, New York.
Davison, A.C. & Hinkley, D.V. (1998) *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.
Darvasi, A. (1997) Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics*, **18**, 19–24.
Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistics Society, Series B*, **39**, 1–38.
Devlin, B. & Risch, N. (1996). Disequilibrium mapping composite likelihood for pairwise disequilibrium. *Genomics*, **36**, 1–16.
El-Assal, S.E., Alonso-Blanco, C., Peeters, A.J.M., Raz, V. & Koorneef, M. (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genetics*, **29**, 435–440.
Falconer, D.S. & Mackay, T. (1996) *Introduction to Quantitative Genetics*, 4th edition. Longman, UK.
Fisher, R.A. (1918) The correlations between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society, Edinburgh*, **52**, 399–433.
Frary, A., Nesbitt, T.C., Frary, A., *et al*. (2000) FW2.2. A quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.
Fulker, D.W. & Cardon, L.R. (1994) A sib-pair approach to interval mapping of quantitative trait loci. *American Journal of Human Genetics*, **54**, 1092–1103.
Gelderman, H. (1975) Investigation on inheritance of quantitative characters in animals by gene markers. *Theoretical and Applied Genetics*, **46,** 300–319.
Goldstein, D.B. & Schlotterer, C. (1999) *Microsatellites: Evolution and Applications.* Oxford University Press.
Guo, S.W. & Thompson, E. (1992) A Monte Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics*, **51**, 1111–1126.
Haley, C.S. & Knott, S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.
Hamrick, J.L. & Godt, M.J.W. (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds A.H.D. Brown, M.T. Clegg, A.L. Kahler & B.S Weir), pp. 43–63. Sinauer, Sunderland, MA.
Hastbacka, J., de la Chapelle, A., Kaitila, I., *et al.* (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics*, **2**, 204–211.
Hastbacka, J., de la Chapelle, A., Mahtani, M.M., *et al.* (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell*, **78**, 1073–1087.
Howell, P.M., Marshall, D.F. & Lydiate, D.J. (1996) Towards developing inter-varietal substitution lines in *Brassica napus* using marker-assisted selection. *Genome*, **39**, 348–358.
Hyne, V., Kearsey, M.J., Pike, D.J. & Snape, J.W. (1995) QTL analysis: unreliability and bias in estimation procedures. *Molecular Breeding*, **1**, 273–282.
Jansen, R.C. & Stam, P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, **136**, 1447–1455.
Kamolsukyunyong, W., Ruanjaichon, V., Siangliw, M., *et al.* (2001) Mapping of quantitative trait locus related to submergence tolerance in rice with aid of chromosome walking. *DNA Research*, **8**, 163–171.
Kaplan, N.L. & Weir, B.S. (1997) The use of linkage disequilibrium for estimating the recombination fraction between a marker and a disease gene. In: *Progress in Population and Human Evolution* (eds P. Donnelly & S. Tavare), pp. 207–219. Spring-Verlag, New York.
Kaplan, N.L., Hill, W.G. & Weir, B.S. (1995) Likelihood methods for locating disease genes in nonequilibrium populations. *American Journal of Human Genetics*, **56**, 18–32.

Kao, C.H. & Zeng, Z.B. (1997) General formulae for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics*, **53**, 653–665.

Kao, C.H., Zeng, Z.B. & Teasdale, R. (1999) Multiple interval mapping for quantitative trait loci. *Genetics*, **140**, 1111–1127.

Kearsey, M.J. (2002) QTL analysis: problems and (possible) solutions. *Plant Journal*, **31**, 355–364.

Kearsey, M.J. & Farquhar, A.G.L. (1998) QTL analysis in plants; where are we now? *Heredity*, **80**, 137–142.

Kearsey, M.J. & Pooni, H.S. (1996) *The Genetical Analysis of Quantitative Traits*. Chapman & Hall, London.

Koumproglou, R., Wilkes, T.M., Townson, P., *et al*. (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *Plant Journal*, **31**, 355–364.

Lagercrantz, U., Putterill, G., Coupland, G. & Lydiate, D. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant Journal*, **9**, 13–20.

Lander, E.S. & Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.

Lander, E.S. & Schork, N.J. (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.

Law, C.N. (1967) The location of genetic factors controlling a number of quantitative characters in wheat. *Genetics*, **56**, 445–461.

Law, C.N., Snape, J.W. & Worland, A.J. (1983) Aneuploidy in wheat and its use in genetical analysis. In: *Wheat Breeding and its Scientific Basis* (ed. F.G.H. Lupton), pp. 71–108. Chapman & Hall, London.

Lehmann, E.C. (1986) *Testing Statistical Hypotheses*. John Wiley & Sons, New York.

Lin, H.X., Yamamoto, T., Sasaki, T. & Yan, M. (2000) Characterisation and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. *Theoretical and Applied Genetics*, **101**,1021–1028.

Luo, Z.W. (1998) Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity*, **80**, 198–208.

Luo, Z.W. & Kearsey, M.J. (1992) Interval mapping of quantitative trait loci in an F_2 population. *Heredity*, **69**, 236–242.

Luo, Z.W. & Suhai, S. (1999) Estimating linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Genetics*, **151**, 359–371.

Luo, Z.W. & Wu, C.I. (2001) Modeling linkage disequilibrium between a polymorphic marker locus and a locus affecting complex dichotomous traits in natural populations. *Genetics*, **158**, 1785–1800.

Luo, Z.W., Tao, S.H. & Zeng, Z.B. (2000) Inferring linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Genetics*, **151**, 457–467.

Lynch, M. & Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, MA.

Mackay, T.F.C. (2001a) Quantitative trait loci in *Drosophila*. *Nature Review Genetics*, **11**, 11–20.

Mackay, T.F.C. (2001b) The genetic architecture of quantitative trait. *Annual Reviews of Genetics*, **35**, 303–339.

Mangin, B., Goffinet, B. & Rebei, A. (1994) Constructing confidence intervals for QTL location. *Genetics*, **138**, 1301–1308.

Manly, K.F. & Olson, J.M. (1999) Overview of QTL mapping software and introduction to map manager QT. *Mammalian Genome*, **10**, 327–224.

Mather, K. (1941) Variation and selection of polygenic characters. *Journal of Genetics*, **41**, 159–193.

Mather, K. & Jinks, J.L. (1982) *Biometrical Genetics.* Chapman & Hall, London.
Melchinger, A.E., Utz, H.F. & Schön, C.C. (1998) QTL mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, **149**, 383–403.
Melchinger, A.E., Utz, H.F. & Schön, C.C. (2000) From Mendel to Fisher: The power and limits of QTL mapping for quantitative traits. *Vortr. Pflanzenzüchtg.*, **48**, 132–142.
Monforte, A.J. & Tanksley, S.D. (2001a) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background; a tool for gene mapping and gene discovery. *Genome*, **43**, 803–813.
Monforte, A.J., Friedman, E., Zamir, D. & Tanksley, S.D. (2001b) Comparison of a set of QTL NILs for chromosome 2 of tomato: Deductions about natural variation and implications for germplasm utilization. *Theoretical and Applied Genetics*, **102**, 572–590.
Morgante, M. & Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant Journal*, **3**, 175–182.
Nilsson-Ehle, H. (1909) Kreuzunguntersuchungen an Hafer und Weizen, Lund.
Rae, A.M., Howell, E.C. & Kearsey, M.J. (1999) More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. *Heredity*, **83**, 586–596.
Ramsay, L.D., Jennings, D.E., Bohuon, E.J.R., *et al.* (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. *Genome*, **39,** 558–567.
Risch, N. & Zhang, H. (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*, **268**, 1584–1589.
Schafer, A.J. & Hawkins, J.R. (1998) DNA variation and the future of human genetics. *Nature Biotechnology*, **16**, 33–39.
Sears, E.R. (1953) Nullisomic analysis in common wheat. *American Naturalist*, **87**, 245–253.
Staub, J.E., Serquen, F.C. & Gupta, M. (1996) Genetic markers, map construction, and their application in plant breeding. *HortScience*, **31**, 729–741.
Tanksley, S.D. (1993) Mapping polygenes. *Annual Reviews of Genetics*, **27**, 205–233.
Tanksley, S.D. & Nelson, J.C. (1996) Advanced backcross analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, **92**, 912–203.
Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. (1989) RFLP mapping in plant breeding: new tools for an old science. *BioTechnology*, **7**, 257–264.
Terwilliger, J.D. (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait locus and one or more polymorphic marker loci. *American Journal of Human Genetics*, **60**, 197–204.
Thoday, J.M. (1961) Location of polygenes. *Nature*, **191**, 368–370.
Thoday, J.M. (1979) Polygene mapping: Uses and limitations. In: *Quantitative Genetic Variation* (eds J.N. Thompson & and J.M. Thoday), pp. 219–233. Academic Press, New York.
Thompson, E.A. & Neel, J.V. (1996). Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. *American Journal of Human Genetics*, **60**, 197–204.
Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovic, S., Nielsen, D. & Buckler, E.S. (2001) Dwarf 8 polymorphisms associated with variation in flowering time. *Nature Genetics*, **28**, 286–289.
Utz, H.F., Melchinger, A.E. & Schön, C.C. (2000) Bias and sampling error of the estimated proportion of genetic variance explained by QTL determined from experimental data in maize using cross validation and validation with independent samples. *Genetics*, **154**, 1839–1849.
van Berloo, R., Aalbers, H., Werkman, A. & Niks, R.E. (2001) Resistance QTL confirmed through the development of QTL NILs for barley leaf rust resistance. *Molecular Breeding*, **8**, 187–195.

Visscher, P.M., Thompson, R. & Haley, C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics*, **143**, 1013–1020.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. & Hornes, M. (1997) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Westman, A.L. & Kresovich, S. (1997) Use of molecular marker techniques for description of plant genetic variation. In: *Biotechnology and Plant Genetic Resources: Conservation and Use* (eds J.A. Callow, B.V. Ford-Lloyd & H.J. Newbury). CAB International, Wallingford, UK.

Wolfe, K.H., Gouy, M.L., Yang, Y.W., Sharp, P.M. & Li, W.H. (1989) Date of the monocot dicot divergence estimated from chloroplast DNA-sequence data. *Proceedings of the National Academy of Sciences, USA*, **86**, 6201–6205.

Xiong, M. & Guo, S.W. (1997) Fine-scale genetic mapping based on linkage disequilibrium: theory and application. *American Journal of Human Genetics*, **60**, 1513–1531.

Xu, S. (1997) A comment on the simple regression method for interval mapping. *Genetics*, **141**, 1189–1197.

Yano, M., Katayose, Y., Asikari, M., *et al*. (2000) *HD1*, a major photoperiod sensitivity trait locus in rice, is closely related to *Arabidopsis* flowering time gene, *CONSTANS*. *Plant Cell*, **12**, 2473–2483.

Zeng, Z.B. (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proceedings of the National Academy of Sciences, USA*, **90**, 10972–10976.

Zeng, Z.B. (1994) Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457–1468.

2 Marker-assisted breeding

Frédéric Hospital

2.1 Introduction

In contrast to past decades, when almost no markers were available and breeding was mostly based on the selection of phenotype, advances in molecular genetics have enabled the partial dissection of the 'black box' of quantitative traits. The use of molecular genetics rests on the ability to determine the genotype of individuals using DNA analysis (Mohan *et al.* 1997; Westman & Kresovich 1997; de Vienne 2002), and results in two types of (molecular) information:

1. identified loci, which are (rarely) causal mutations; and
2. (more frequently) presumed non-functional genetic markers (indirect markers).

Markers can be used simply to assess the parental origin of an anonymous genomic region, or to unravel the genetic architecture of the quantitative trait(s) of interest, based on evidence of empirical associations of marker genotypes with trait phenotype (QTL detection, see Chapter 1). These associations can then be used for selection. Note here that, whereas causative polymorphisms provide direct information about the genotype for QTL, the use of indirect markers for QTL mapping and for selection is based on the existence of linkage disequilibrium (Hartl & Clark 1997) between the marker and the QTL. Marker-assisted selection can also aim at the introduction of transgenes into breeding populations (see Chapter 4).

Molecular information can be used in several ways to make the plant or animal breeding process more efficient, mainly with so-called marker-assisted selection (MAS) schemes (Dekkers & Hospital 2002). Herein, some of the most promising techniques involving selection on molecular markers in order to increase the speed and/or the efficiency of plant breeding programmes will be reviewed; that is, to increase genetic gain per unit of time. Other – possibly important – uses of molecular information that will not be reviewed in this chapter include parentage verification or identification, identification and characterization of genetic resources, and quantification of genetic diversity.

Selection decisions in breeding programmes can be based on phenotypic information alone (conventional selection), on molecular information alone, or on a combination of both. Herein, breeding strategies involving selection based on molecular information alone are termed 'genotype building' (GB) strategies, because

the selection phase can be reduced to a simple 'building blocks' problem. Based on phenotypic and/or molecular information available prior to the start of the selection programme, the breeder defines the ideal genotype (ideotype) at a collection of loci (target loci), as the one that meets the selection objective. The parents originally hosting the different target genes are crossed, after which selection consists of screening – among the different genotypes produced by recombination in one or more generations – the one(s) that is closest to the ideotype, or that allows production of the ideotype most rapidly, based simply on DNA analysis (marker genotypes). Finally, on completion of the MAS programme, phenotypic evaluation is carried out in order to evaluate the agronomic value of the resulting progenies.

One case of GB widely used in plant breeding is marker-assisted introgression in backcross programmes, and this case is reviewed in detail below. Other possible GB schemes (marker-based population screening, recurrent selection or gene pyramiding) are reviewed more briefly. Clearly, phenotypic information may also be used during the selection phase in addition to – or in combination with – molecular information. The corresponding techniques are also reviewed.

Compared to breeding schemes involving phenotypic evaluation at each selection step, GB schemes where selection is based solely on molecular information capitalize on a potentially important saving of time and/or experimental means, particularly in cases where phenotypic information is longer, more difficult, and/or more expensive to score than molecular information (e.g. testcross breeding for yield in maize involving progeny testing, malting quality in barley, breeding for disease resistance in most crops). However, such a breeding strategy assumes that the effects associated with markers are sufficiently well estimated, and sustainable across agronomic conditions and genetic background, so that the realized genetic gain will meet expectations. The risk inherent in this assumption is discussed in the light of some recently published results of GB experiments in plants.

2.2 Marker-assisted backcrossing of a single target gene

Backcross breeding is a well-known procedure for the introgression of a target gene from a donor line into the genomic background of a recipient line. The target locus (chromosomal location of the target gene) is kept heterozygous (donor/recipient) by selection for the donor type allele at each generation. On chromosomal locations outside of the target locus, the objective is to increase the recipient genome content (RGC, often expressed as the percentage of recipient alleles) of the progenies. Increasing RGC is particularly important (but difficult) on the chromosome carrying the target locus (carrier chromosome). Because at each generation there is selection for the donor allele at the target locus on the carrier chromosome, unwanted donor genes located in the vicinity of the target locus may be dragged along with the target gene (linkage drag). Hence, RGC is expected to be lower on the carrier chromosome than on non-carrier chromosomes.

In conventional breeding schemes, presence of the target gene is assessed phenotypically – provided that this gene has a visible or measurable effect – and any increase in RGC is ensured by repeatedly backcrossing progenies carrying the target gene to the recipient line. Given that donor content on non-carrier chromosomes is expected to be reduced by one half after each backcross, it is generally assumed that at least six or seven backcross generations are necessary to ensure a sufficient genomic similarity between the backcross line and the recipient line (except the target locus). In practice, the number of backcrosses performed is sometimes even larger (10 or more). Using molecular information may improve the efficiency of backcross breeding schemes in several non-exclusive ways, through selection on molecular markers (marker-assisted backcrossing) to:

1. control the target gene (foreground selection); and
2. control the genetic background (background selection).

In all cases, marker assistance is expected to provide higher efficiency, reduced cost, and/or shorter duration of the backcross breeding scheme, compared with conventional methods. In addition, markers can also be used to estimate RGC in the backcross progenies. Such uses of markers are detailed below in the case of a backcross programme involving one single target gene. Additional considerations related to the manipulation of more than one gene are addressed in the following section, along with other marker-assisted breeding strategies for several genes.

Marker-assisted backcrossing is of great practical interest in applied breeding schemes, either to manipulate 'classical' genes between elite lines or from genetic resources, or to manipulate transgenic constructions, or quantitative trait loci. From a theoretical standpoint, this is a 'simple' example of marker-based selection: in general, only two alleles are segregating, and the gametic phase (parental origin of the alleles on a chromosome) is known because recombination is effective only for one chromosome of each pair (the chromosome from the gamete produced by the backcrossed parent). It is then an appropriate case study to investigate how selection and recombination work together to make it work better in any type of MAS programme.

2.2.1 Foreground selection

Here, we address the use of molecular markers to assess the presence of the target gene in backcross progenies.

2.2.1.1 Target locus is a known locus

Molecular data can be obtained at an early stage as soon as DNA can be obtained (e.g. from a leaf sample). Heritability at the molecular marker level equals one, and most often it is possible to find markers for which the dominance relationship is favourable. Conversely, phenotypic assay is often longer and/or more difficult and/or more costly than molecular genotyping. Classical examples are cases when phenotypic

assay involves progeny testing, because the phenotype is expressed after reproduction (e.g. grain yield, any testcross performance, malting quality in barley), because the phenotypic assay is destructive, or because the target gene is recessive. In such circumstances, even when the target is a major gene (of known phenotypic effect and chromosomal location), control of the target with molecular markers may be more profitable than phenotypic assay.

When the presence of the target gene is not controlled directly through its phenotypic expression, but indirectly through the observed genotype at one or several marker(s), it is important that observed genotype at the marker provides a good control over the (true but unknown) genotype at the target. This may be assessed, for example, by computing the target control rate (TCR) defined for any given individual as the probability that the individual is heterozygous donor/recipient at the target given that it is heterozygous donor/recipient at the marker. If we denote the donor allele at the target as T^D, T^R as the recipient allele at the target, M^D as the donor allele at the marker and M^R as the recipient allele at the marker, we have:

$$\text{TCR}\,(\%) = \Pr\{(T^D/T^R) \mid (M^D/M^R)\} \times 100$$

$$= \Pr\{(T^D M^D/T^R M^R)\}/\Pr\{(M^D/M^R)\} \times 100$$

where Pr{(X/Y)} denotes the probability of being of genotype X/Y, and | denotes condition. In other words, the risk that an individual, which displays the desired genotype at the marker(s), does *not* have the desired genotype at the target is (100 – TCR).

In some favourable cases, it is possible to find a direct marker inside the target locus. An obvious example is when the target gene is a transgenic construct of which the complete DNA sequence is known. In such cases, the recombination rate between the marker and the target locus is zero, the probability of transmission of the marker is 50% at each backcross (BC) generation, as is the probability of transmission of the target gene, and the genotype at the target locus equals that at the marker, so TCR = 100%. However, such cases are rare. In general, the target has to be controlled by indirect markers located outside of the target locus, so that the recombination rate between the target and the marker(s) is not zero.

If the target is controlled by one single marker M_1 such that the recombination rate between the target and the marker is r_1, then after n BC generations:

$$\text{TCR}(M_1) = \{(1/2)^n (1 - r_1)^n\}/\{(1/2)^n\} = (1 - r_1)^n$$

If the target is controlled by two markers M_1 and M_2 (one on each 'side' of target locus T) such that the recombination rate between T and M_1 is r_1, the recombination rate between T and M_2 is r_2, and the recombination rate between M_1 and M_2 is r, then after n BC generations:

$$TCR(M_1M_2) = \{(1/2)^n (1 - r_1)^n (1 - r_2)^n\}/\{(1/2)^n (1 - r)^n\}$$
$$= (1 - r_1)^n (1 - r_2)^n/(1 - r)^n$$

Tabulated numerical values for $TCR(M_1)$ or $TCR(M_1M_2)$ are given in Table 2.1 for a range of marker positions, where marker position is expressed as the distance between the target and the marker in Haldane centimorgans (cM), assuming no interference in recombination. It is clearly seen from Table 2.1 that control of the target by a single marker is not satisfactory in most cases. The marker must be as close as 1 cM to the target to keep the risk of 'losing' the target below 5% over five BC generations. Even with a single marker at 1 cM, the risk of losing the target is close to 10% in BC_{10}. For greater distances of a single marker, the risk rapidly becomes too high. Conversely,

Table 2.1 Target control ratio (TCR) and minimal population size when the target is a known gene.

(a) Target controlled by single marker

Target–marker distance (cM)	TCR (%)					
	BC1	BC2	BC3	BC5	BC10	N_{min}
1	99.0	98.0	97.1	95.1	90.5	7
3	97.1	94.3	91.5	86.3	74.4	7
5	95.2	90.7	86.4	78.4	61.4	7
10	90.9	82.7	75.2	62.2	38.7	7
15	87.0	75.8	65.9	50.0	25.0	7
20	83.5	69.7	58.3	40.6	16.5	7
25	80.3	64.5	51.8	33.4	11.2	7
30	77.4	60.0	46.4	27.9	7.8	7

(b) Target controlled by marker bracket

Target–marker distance (cM)	TCR (%)					
	BC1	BC2	BC3	BC5	BC10	N_{min}
1	100.0	100.0	100.0	100.0	99.9	7
3	99.9	99.8	99.7	99.6	99.1	8
5	99.8	99.5	99.3	98.8	97.5	8
10	99.0	98.0	97.1	95.2	90.6	9
15	97.8	95.7	93.6	89.6	80.3	10
20	96.3	92.6	89.2	82.6	68.2	11
25	94.3	89.0	84.0	74.7	55.8	12
30	92.2	85.0	78.3	66.5	44.3	12

TCR(%) is the probability that an individual at the given BC generation has the desired genotype at the target given that it has the desired genotype at the flanking marker(s). N_{min} is the minimum number of individuals that should be genotyped at each BC generation to obtain at least one individual with the desired genotype at the flanking marker(s) with risk $\alpha_M = 1\%$. See text for details.

the data in Table 2.1 show that control of the target by two markers (marker bracket) is much more satisfactory, even for larger marker distances. A bracket with each marker as far as 10 cM from the target provides approximately the same control as a single marker at 1 cM, and the control provided by any bracket closer than 10 cM is quite satisfactory. Control provided by brackets larger than 10 cM is also acceptable, but only for a few BC generations. Obviously, this is because when the target is controlled by two markers a (rare) double recombination is necessary to break the bracket-target association, while a (more frequent) single recombination is sufficient to break a single marker-target association.

2.2.1.2 Target locus is a quantitative trait locus (QTL)

A QTL is a locus, or a chromosomal segment involved in the variability of a quantitative trait, which is detected by appropriate statistical methods that correlate molecular and phenotypic information (see Chapter 1). By definition, such a target can only be controlled by indirect markers (unless it is characterized at the molecular level, which is rarely the case and necessitates a lot more additional work). The control of a target QTL in foreground selection poses additional problems, because the exact location of the target is often not known, but rather estimated with a given precision. Hence, the number and chromosomal positions of the markers devoted to target control must take account of the uncertainty of the true target location. In extending a preliminary result reported by Visscher *et al.* (1996), Hospital and Charcosset (1997) discussed the optimal number and positions of markers to control a QTL in the foreground selection step of a BC programme. Calculations are based on the following rationale: it is assumed that QTL detection has already been performed, and has provided the expected (most likely) position of the QTL, along with a confidence interval on that position. Then, uncertainty in the target position is modelled by assuming that the true target is located somewhere around the expected QTL position, with a distribution following a normal law of mean 0 and of variance derived from the length of the confidence interval. For a given number of markers, Hospital and Charcosset (1997) then computed the optimal positions of the markers as the ones that maximize the TCR. For each possible position of the true target, TCR is computed (as above) for a known target, and then averaged over all possible positions. Some relevant numerical results are given in Table 2.2.

It is seen that target control provided by optimally placed markers is very good in BC_1 and BC_3. In general, three markers optimally placed are sufficient to ensure a TCR above 99%, even for large confidence intervals (60 cM). Even fewer markers can be used to control smaller confidence intervals. However, Hospital and Charcosset (1997) showed that control by few markers (three or less) is quite sensitive to marker positions, so that in the normal situation where markers cannot be placed optimally, using at least three markers per QTL is recommended, except when precision on QTL location is very high.

Table 2.2 Target control ratio (TCR) and minimal population size when the target is a QTL.

Confidence interval length (cM)	Number of markers	Optimal marker positions (distance from marker to expected QTL position, cM)					TCR (%)		
							BC1	BC3	N_{min}
10	1	0.0					98.5	95.6	7
	2	–3.6	+3.6				99.9	99.6	8
	3	–4.7	0.0	+4.7			100.0	99.9	8
	4	–5.4	–1.4	+1.4	+5.4		100.0	99.9	8
	5	–5.9	–2.2	0.0	+2.2	+5.9	100.0	100.0	8
20	1	0.0					97.0	91.5	7
	2	–6.2	+6.2				99.6	98.7	8
	3	–8.5	0.0	+8.5			99.8	99.5	9
	4	–9.9	–2.6	+2.6	+9.9		99.9	99.8	9
	5	–10.8	–4.1	0.0	+4.1	+10.8	100.0	99.9	9
40	1	0.0					94.4	84.5	7
	2	–10.4	+10.4				98.7	96.4	9
	3	–14.9	0.0	+14.9			99.5	98.5	10
	4	–17.7	–4.8	+4.8	+17.7		99.7	99.2	11
	5	–19.6	–7.8	0.0	+7.8	+19.6	99.8	99.5	11
60	1	0.0					91.9	78.5	7
	2	–14.0	+14.0				97.8	93.6	10
	3	–20.5	0.0	+20.5			99.0	97.1	11
	4	–24.6	–6.8	+6.8	+24.6		99.5	98.4	13
	5	–27.5	–11.2	0.0	+11.2	+27.5	99.7	99.0	13

TCR(%) is the probability that an individual at the given BC generation has the desired genotype at the target given that it has the desired genotype at the flanking marker(s). N_{min} is the minimum number of individuals that should be genotyped at each BC generation to obtain at least one individual with the desired genotype at the flanking marker(s) with risk $\alpha_M = 1\%$. See text for details.

2.2.1.3 Minimal population sizes

When the target is controlled indirectly via molecular markers, it is never possible to know whether an individual that carries the desired genotype at foreground selection markers (i.e. an individual that is heterozygous at all markers devoted to target control) also really carries the desired genotype at the target locus (i.e. is heterozygous at that locus) until a phenotypic assay is performed. Hence, a strategy could be to:

1. choose marker numbers and positions based on the above calculations so that these markers provide as high as possible TCR;
2. then compute minimal population sizes such that the risk of not obtaining at least one individual with the desired genotype at the markers is below a given threshold, say α_M.

With this strategy, the risk of not obtaining an individual with the desired genotype at the target is simply $(100 - TCR) \times \alpha_M$, which supposedly is low enough. Computing minimal population sizes in this context is quite simple.

At each BC generation the probability that an individual has the desired marker genotype given that his backcrossed parent had the desired genotype is:

$$P_M = ½ (1 - r_1) \times (1 - r_2) \times \ldots \times (1 - r_{m-1})$$

where m is the total number of markers and r_k is the recombination rate between the k^{th} and $(k + 1)^{th}$ markers. The minimum number N_{min} of individuals that should be genotyped at each BC generation to obtain at least one individual with the desired genotype at the markers with risk α_M is obtained by solving the equation $(1 - P_M)^N = \alpha_M$, so that:

$$N_{min} = \ln (\alpha_M)/\ln (1 - P_M)$$

where ln denotes the Naperian logarithm. Numerical values for N_{min} are given in the last columns of Tables 2.1 and 2.2 for the corresponding marker positions. Minimal population sizes are quite low (between 7 and 13 individuals at each BC generation) in all cases, even for QTL located with poor precision. However, the following should be borne in mind.

1. This ensures that the genotype at *flanking markers* is obtained, and one should check that the corresponding TCR allows a sufficient control of the target.
2. This ensures that at least *one* individual with desired genotype is obtained; more complex calculations to obtain several carriers of the target gene were derived by Melchinger (1990).
3. This is for the control of a single target (either known gene or QTL).

For BC breeding programmes aiming at introgression of several targets, minimal population sizes rapidly increase with the number of targets, and turn out to be one of the most limiting parameters. This is addressed later in the chapter (see Section 2.3).

2.2.2 Background selection

Whether the target is controlled directly through its phenotypic expression, or indirectly by markers (foreground selection, see Section 2.2.1.3), molecular markers can always be used for background selection. The aim of background selection is to hasten the recovery of recipient parent genotype outside the target(s) (genetic background). Usually, foreground and background selection are performed in two distinct and successive steps at each BC generation, because it is not necessary to genotype the background of individuals that do not carry the target (unless such global genotyping is less expensive because of the particular molecular technique used).

In any case, it is assumed here that selection is in two steps, and that step 1 (foreground selection) can be achieved successfully, either using phenotype or markers,

so that the TCR is close to 100%, and step 2 (background selection) is amongst individuals that are all heterozygous for the target.

Even when no background selection is performed (random selection), the percentage of alleles inherited from the recipient parent (RGC) is expected to increase in BC progenies, due to successive backcrosses. Hence, background selection is considered efficient only if it permits a return to the recipient genome faster than the normal return rate when no selection on markers is applied; the efficiency of MAS should always be compared to this normal rate as a reference. We need first to recall what is the normal rate of return expected with no background selection.

2.2.2.1 Expected genome contents with no selection

On non-carrier chromosomes (chromosomes not hosting the target), the probability that any locus remains heterozygous donor/recipient after n backcrosses is $(½)^n$. Hence, starting from 50% in the original F_1, the expected RGC(%) on non-carrier chromosomes at generation BC_n with no background selection is $100 \times \{1 - (1/2)^{(n+1)}\}$. Based on this consideration, it is generally considered that at least six BC generations are necessary to ensure a similarity with the recipient parent above 99%. However, it is important to note that on a carrier chromosome (chromosome hosting the target), the rate of return to recipient type is much slower than that.

On the carrier chromosome, selection of individuals that are heterozygous at the target locus is mandatory. Hence, because of this foreground selection, any locus that is linked to the target locus on the carrier chromosome is more likely to be heterozygous than a locus on a non-carrier chromosome. This linkage drag makes the RGC on the carrier chromosome with no background selection always lower than the RGC on the non-carrier chromosome. There are two possible measures of linkage drag:

1. One can compute the *total* proportion of donor alleles on the carrier chromosome. This was provided by Stam and Zeven (1981) based on the following calculation. Given that the target locus T is heterozygous, any locus X on the carrier chromosome is heterozygous with probability (1 – r), where r is the recombination rate between T and X. Then, the expected proportion of loci that are heterozygous is obtained by integrating along the chromosome, assuming no interference in recombination: i.e. the relationship between map distance d (in cM) and corresponding recombination rate r[d] is obtained from Haldane's formula: r[d] = ½ (1 – Exp[–2d/100]), where Exp[] denotes exponential.
2. Also, one can focus on the intact donor segment; that is, the chromosome segment of donor origin containing the target locus, which has remained unaltered by crossovers since the original cross between the donor and recipient parents. Hanson (1959) first provided the theoretical expression for the expected length of this intact segment. This was revisited by Naveira and Barbadilla (1992), who also provided the corresponding variance. The computation similarly involves integration along the chromosome, but here the relevant probability is that of absence of crossover between T and X, not absence of recombination as in the previous case. Recall that the probability of absence of recombination

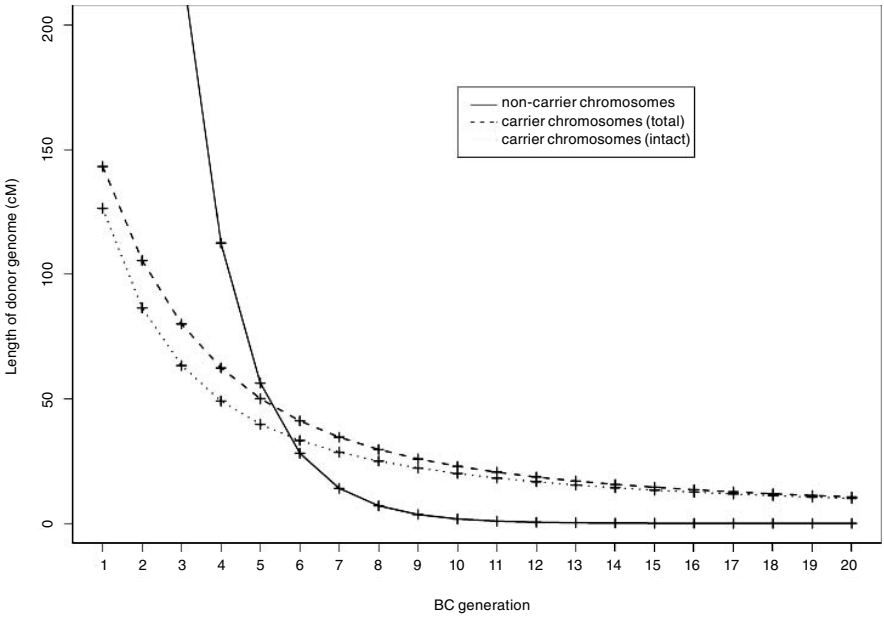


Figure 2.1 Donor genome contents. Expected length of chromosome segments of donor origin (ordinate, in Haldane centimorgans, cM) still segregating in the population at different backcross (BC) generations (abscissa) on the different parts of the genome: non-carrier chromosomes (solid line), whole-carrier chromosome (dashed line) or the intact segment of donor origin on the carrier chromosome (dotted line) (see text for more detail). The genome considered here consists of one carrier chromosome and nine non-carrier chromosomes, each of 200 cM length.

between two loci is the probability of an even number (0, 2, 4, …) of crossovers between the two loci, not only zero crossovers. Clearly, the *total* proportion of donor alleles on the carrier chromosome computed by Stam and Zeven (1981) comprises the length of the intact donor segment plus other blocks of donor alleles elsewhere on the carrier chromosome.

The quantity of donor genes (in cM) on the different parts of the genome is shown in Figure 2.1 for a genome composed of one carrier chromosome and nine non-carrier chromosomes, each chromosome of length 200 cM. Because there are nine non-carrier chromosomes, most of the unwanted donor genes are located on non-carrier chromosomes in early BC generations, but these genes are rapidly removed as shown. Conversely, the quantity of donor genes on the carrier chromosome decreases much more slowly, so that after generation BC_6 most of the unwanted donor genes still segregating are located on the carrier chromosome. Donor fragments on the carrier chromosome represent a total of 40 cM in BC_6, 23 cM in BC_{10}, and still more than 10 cM in BC_{20}. Although these fragments represent only 1% of the total genome length (4000 cM) after generation BC_6, these may still host numerous unwanted genes, in particular if the donor is a wild genetic resource. Moreover, the

variance around these expected values is important (Stam & Zeven 1981; Naveira & Barbadilla 1992). It is also clear from Figure 2.1 that the difference between the two measures of linkage drag on the carrier chromosome is small. Hence, the vast majority of unwanted donor genes on the carrier chromosome are located on the intact donor segment surrounding the target. An impressive experimental proof of this was provided by Young and Tanksley (1989a) who genotyped *a posteriori* with RFLPs a collection of tomato varieties previously introgressed for the resistance gene at the *Tm-2* locus. The size of chromosomal segments retained around the *Tm-2* locus during backcross breeding was very variable and sometimes quite long: one line exhibited a donor segment of 50 cM after 11 backcrosses, another 36 cM after 21 backcrosses, etc.

2.2.2.2 Marker-based estimate of recipient genome content

In a particular experiment, it is possible to use molecular markers simply to estimate the RGC of backcross progenies. The most basic estimate is provided by scoring the genotype at a collection of markers over the genome, and then estimating RGC from the ratio of the number of markers homozygous for the recipient allele over the total number of markers scored. This simplest estimate does not take into account the positions of the markers, but it is self evident that if the markers are not evenly distributed along the genome (the real situation), then weighting them equally is clearly not the best solution. Several solutions have been proposed to take account of markers locations. Visscher (1996) proposed including molecular information in a BLUP-like estimate of RGC, and derived the optimal positions of markers in this context. Young and Tanksley (1989b) introduced the concept of graphical genotypes to 'portray the parental origin and allelic composition throughout the genome'. For each chromosomal segment flanked by two markers, RGC is approximated based on the genotypes of the flanking markers: 100% if two markers of recipient type; 0% if two markers of donor type; and 50% if one marker of donor type and one marker of recipient type. This ignores the possible recombination events taking place between the two flanking markers. To take recombination into account, one can compute for any point of a chromosome the probability of being of recipient type, given marker genotypes, marker positions, and the breeding scheme (Servin *et al.* 2002). In any case, the general conclusion is that few well-placed markers (two to four markers on a chromosome of 100 cM) provide adequate coverage of the genome in backcross programs (Visscher 1996; Servin & Hospital 2002).

2.2.2.3 Reduction of linkage drag (carrier chromosome)

As seen above (Section 2.2.2.1), the carrier chromosome deserves special consideration in backcross programmes because this chromosome returns to the recipient type more slowly than non-carrier chromosomes, due to selection for the target gene in each generation (linkage drag). The following text will focus only on the intact donor segment (see above) as a measure of linkage drag.

Basically, linkage drag can be reduced by performing background selection at two markers flanking the target, one on each side. The objective is to select individu-

als that are heterozygous at the target locus, and homozygous for the recipient allele at both flanking markers (such individuals are termed double recombinants).

Hospital (2001) computed the mean and variance of the length of the intact donor segment around the target gene, for double-recombinant individuals, in any BC generation. This gives the efficacy of background selection for the reduction of linkage drag. The numerical results indicate that the expected length of donor segment on each side of the target gene is approximately half the distance between the target and the flanking marker in BC_1. The length in more advanced BC generations depends on the marker distance, but for short marker distances (20 cM from the target or less), the expected length of donor segment in advanced BC generations is not much below the length in BC_1. For short marker distances, recombination events are rare and do not accumulate; in general, the genotypes selected experienced only one crossover – the one that permitted the flanking marker to return to the recipient genotype. The basic conclusion is that selecting for distant markers over several successive backcross generations cannot provide a better reduction of linkage drag than using close markers. Using very close markers is the only way to make any substantial reduction in linkage drag.

Clearly, selecting for flanking markers close to the target implies genotyping and screening large populations before a double-recombinant genotype is obtained. In order to optimize genotyping effort (i.e. the cost of the programme), it is therefore important to determine the minimal population sizes necessary to obtain the desired genotypes at the flanking markers. Intuitively, for close flanking markers, double-recombinant genotypes are highly unlikely to be obtained in one single generation (BC_1), so that at least two BC generations should be performed, with selection for a single recombinant genotype on one side of the target in BC_1, and a single recombinant on the other side in BC_2 (Young & Tanksley 1989a). The underlying mathematics have been elucidated only recently, however. A first solution was derived by Hospital and Charcosset (1997), and this result was used by Frisch *et al.* (1999) with numerical applications in the context of single-generation optimization (assuming that the genotype selected at generation $BC_{(n)}$ is known, population size at generation $BC_{(n+1)}$ is optimized to permit the selection of a double recombinant genotype at generation $BC_{(n+1)}$). However, Hospital (2001) showed that a better optimization is obtained when considering all the planned generations simultaneously, because the optimal population size at each BC generation depends on the total duration of the breeding scheme.

Optimizing population sizes over several successive generations requires some numerical calculations. A computer program (*popmin*) that performs these calculations easily was designed (Hospital & Decoux 2002) and is freely available at http://moulon.inra.fr/~fred/programs. This program works as follows. The user enters a given value n for the total duration of the breeding scheme (maximal number of BC generations that could be performed) and a given risk, α. Consider a marker-assisted backcross scheme involving n generations with populations sizes $N_{1...}N_n$ at generations $BC_{1...}BC_n$, respectively. The selection objective is here to obtain a double-recombinant genotype at any generation BC_k ($k \leq n$) but, obviously, if a double

recombinant is obtained at generation $k < n$, then the BC scheme is interrupted. The program computes the probabilities, S_k, that a double recombinant is obtained at any generation BC_k ($k \leq n$). Then, it determines optimal populations sizes $N_1 \ldots N_n$ at generations $BC_1 \ldots BC_n$ such that: (i) the risk that *no* double recombinant is obtained after n BC generations is α ($\Sigma_k S_k \geq 1 - \alpha$); and (ii) the average number of individuals genotyped ($N^* = \Sigma_k N_k S_k$) is minimal.

The *popmin* program can be run to investigate any particular situation, but the general conclusions that should be borne in mind are as follows. First, it is often preferable to genotype *more* individuals in advanced BC generations than in early BC generations (e.g. for a BC scheme lasting two generations, genotype more individuals in BC_2 than in BC_1, not the reverse). This reduces the average number of genotypings over the entire BC scheme. Second, planning to perform a total of more than two BC generations is in general recommended.

This is exemplified in Table 2.3, which shows numerical results obtained with the *popmin* program, for flanking markers located at 2 cM from the target on each side, and BC schemes of different durations, with a risk $\alpha = 1\%$. The minimum number of individuals that should be genotyped to obtain a double recombinant in BC_1 is about 24 000 – clearly far too many. The same result can be obtained over two generations (BC2 strategy) by genotyping 290 individuals in BC_1, and 499 in BC_2. Finally, over three generations (BC3 strategy), the optimal population sizes are 117 individuals in BC_1, 171 in BC_2, and 371 in BC_3. In all three strategies, the probability of obtaining a double recombinant for the flanking markers by the end of the breed-

Table 2.3 Example of optimal populations sizes computed by the *popmin* program.

Maximal duration (n)	Generation (BC_k)	Population size (N_k)	Cumulated genotypings ($G_k = \Sigma_{i \leq k} N_i$)	Probability of success (S_k)	Cumulated probability ($\Sigma_{i \leq k} S_i$)	Averaged genotypings ($\Sigma_{i \leq k} G_i S_i$) / ($\Sigma_{i \leq k} S_i$)
1	BC_1	23 961	23 961	0.99	**0.99**	**23 961**
2	BC_1	290	290	0.05	0.05	290
	BC_2	499	789	0.94	**0.99**	**764**
3	BC_1	117	117	0.02	0.02	117
	BC_2	171	288	0.72	0.74	283
	BC_3	370	658	0.25	**0.99**	**378**
4	BC_1	72	72	0.01	0.01	72
	BC_2	95	167	0.46	0.47	165
	BC_3	143	310	0.37	0.84	229
	BC_4	303	613	0.15	**0.99**	**287**
5	BC_1	52	52	0.01	0.01	52
	BC_2	64	116	0.30	0.31	114
	BC_3	86	202	0.34	0.65	160
	BC_4	130	332	0.24	0.89	206
	BC_5	259	591	0.10	**0.99**	**245**

Selection for the reduction of linkage drag with two markers located at 2 cM on each side of the target. BC schemes with different total durations n, each with a global risk of $\alpha = 1\%$. The probability of success is the probability to obtain a double recombinant genotype at the given generation. See text for details.

ing scheme is above 99%. In the BC3 strategy, the probability of obtaining a double recombinant in BC_2 is about 75%. If this happens, the BC scheme is obviously not pursued to BC_3 (unless for other reasons not considered here). Hence, planning to perform a maximum of three BC generations (BC3 strategy) permits, in 75% of the cases, the production of a double recombinant in BC_2 by genotyping a total of only 290 individuals, which is much less than the 789 individuals necessary with the BC2 strategy. With the BC3 strategy, only in 25% of the cases should the programme be really continued until generation BC_3. Hence, averaging over all possibilities, the mean number of individuals that need to be genotyped to obtain a double recombinant with the BC3 strategy is only about 380, to be compared with an average of about 760 with the BC2 strategy. Planning at the beginning of the programme to perform more than two BC generations is then always a better strategy to optimize the costs of genotyping (unless a rapid success is really mandatory). This is equivalent to fixing a not-too-low risk of failure per generation (risk of not obtaining a double recombinant at that generation), in particular in early BC generations, which is converse to what was advocated by Frisch *et al.* (1999).

Obviously, the strategy and number of individuals to be genotyped should be reconsidered at each generation once the actual genotype of the individual selected is known. This is also possible using the computer program *popmin* with a relevant option.

Finally, the best optimization strategy is as follows:

1. Before starting a marker-assisted backcross programme, investigate different scenarios to determine the maximal number of BC generations (n) that could be performed, given available genotyping means, and other possible economic considerations.
2. Start the BC scheme with the optimal population size in BC_1 corresponding to the chosen scenario.
3. Refine the optimization at each following generation, once the genotype of the selected individual is known.

To synthesize the above results and help decision-making when designing a marker-assisted backcross scheme in particular conditions, the *popmin* program was run for a range of parameters to provide the direct relationship between the efficiency of MAS (expected linkage drag at the end of the breeding scheme) and the total number of individuals genotyped for flanking markers during the breeding scheme. The results are shown in Figure 2.2 for breeding schemes involving a total of one, two, three, five or ten BC generations. This can be used as follows. One should define the upper limit for the length of linkage drag that is acceptable at the end of the experiment [e.g. for introgression of a gene from wild genetic material, or for fine-mapping of a QTL, and/or for the derivation of near-isogenic lines (NIL) or congenic lines for the identification and validation of quantitative trait loci, acceptable linkage drag should be much smaller than that for introgression between elite lines]. Also, one should define the upper limit for the total number of individuals that can be geno-

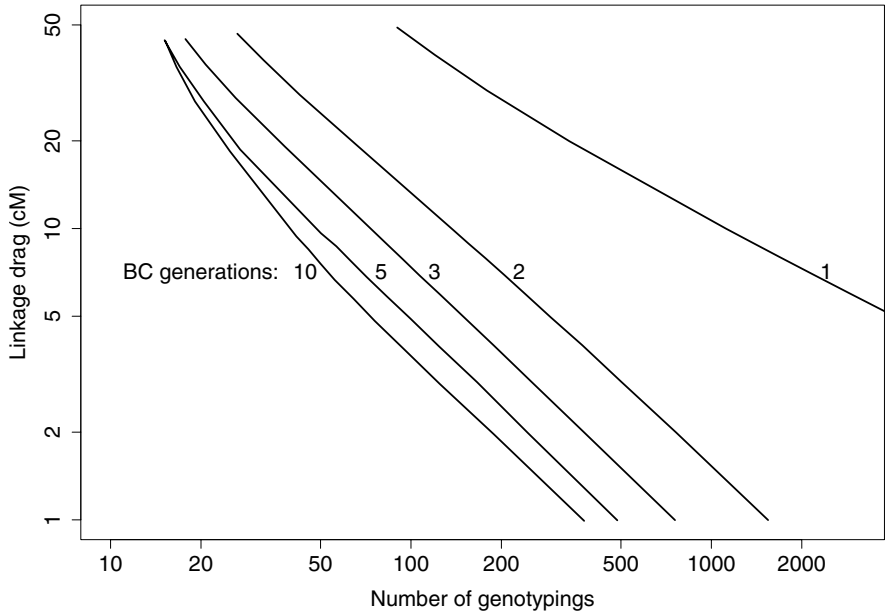


Figure 2.2 Reduction of linkage drag. Relation between the expected length of the chromosome segments of donor origin around the target gene (ordinate, in Haldane centimorgans, cM), and total number of individuals genotyped for the two flanking markers (abscissa). Each line corresponds to marker-assisted backcross (BC) programmes of different durations (number of BC generations indicated in the middle of each line). Note that both abscissa and ordinate are on logarithmic scales.

typed during the experiment (e.g. based on available molecular facilities and cost of the molecular technique used). Figure 2.2 can then be used to determine how many BC generations should be performed in order to remain below the two defined limits. It is seen from the figure that BC schemes involving only one or two BC generations will rarely be affordable, unless large linkage drag can be accepted. Conversely, the small difference between the lines in Figure 2.2 for five versus ten BC generations indicates that performing more than five BC generations is rarely necessary, unless very low molecular cost is sought. Note, however, that both the x (abscissa) and y (ordinate) axis in Figure 2.2 are logarithmic scales, so that small differences on the x-axis may correspond to hundreds of individuals for small linkage drag values.

The general conclusion is then that a typical marker-assisted backcross scheme should involve three to four BC generations in most cases, unless rapid success is sought for particular reasons. Planning to perform three or more BC generations has two main advantages. First, it permits a more drastic reduction of linkage drag while reducing the genotyping effort. Second, it increases the probability of success (obtaining a double recombinant) in advanced BC generations. The optimal population sizes above were defined such that at least one double recombinant is obtained with

a given risk. It is then likely that on average more than one is obtained. Background selection on non-carrier chromosomes is then possible among those double recombinants, as described in the next section.

2.2.2.4 Selection on non-carrier chromosomes

Since the original benchmark reports of Tanksley (1983) and Tanksley *et al.* (1989), many groups have addressed the use of markers to hasten the recovery of recipient genome on non-carrier chromosomes in BC breeding schemes (e.g. Hillel *et al.* 1990; Hospital *et al.* 1992; Groen & Smith 1995; Visscher *et al.* 1996). This was also validated experimentally (see Section 2.6). In such cases, the objective is to select individuals that are of homozygous recipient type at a collection of markers located on non-carrier chromosomes. Again, several markers are involved and it is unlikely that the selection objective is fulfilled in one single generation (BC_1), so that selection on markers should be performed over two or more BC generations.

The general conclusions that can be drawn from these theoretical studies can be summarized as follows. First of all, a dense coverage of the non-carrier chromosomes by molecular markers is not mandatory to increase the overall recipient genome content (unless fine-mapping of particular chromosomal regions is important). For a chromosome of 100 cM, two to four markers are sufficient. Obviously, selection on markers is most efficient if the markers are optimally positioned along the chromosomes. Such optimal positions were derived by Servin and Hospital (2002) and are presented in Table 2.4.

Table 2.4 Optimal positions of markers for background selection on non-carrier chromosomes.

Markers per chromosome	Generation	Optimal marker positions (cM)			RGC (%)	
		d*	d*–	d*+	Selection	No selection
2	BC_1	18.6	10.4	27.0	93.4	75
	BC_2	21.4	10.0	32.8	95.2	87.5
	BC_3	22.9	7.1	38.6	96.9	93.75
3	BC_1	8.4	0	17.9	97.1	75
	BC_2	11.0	0	23.5	97.6	87.5
	BC_3	12.6	0	29.7	98.3	93.75
4	BC_1	4.5	0	14.4	98.5	75
	BC_2	6.5	0	19.5	98.6	87.5
	BC_3	7.8	0	25.2	98.9	93.75

Results are given for a chromosome of length 100 cM, for different numbers of markers per chromosome, at different backcross generations. Optimal marker positions are described by the distance d* (in Haldane centimorgans) from the telomere to the first marker on each chromosome end. Other markers are evenly located on the rest of the chromosome. RGC(%) with selection is the expected recipient genome content for individuals that are homozygous for the recipient allele at all markers. RGC with no selection is recalled for comparison. d*– and d*+ give the range of d values for which RGC is not decreased by more than 1% compared to the value at d*. Data from Servin and Hospital (2002).

However, a precise positioning of the markers on non-carrier chromosomes is again not mandatory (contrary to the case of the reduction of linkage drag on the carrier chromosome, see above). As can be seen from Table 2.4 (d*–/d*+), a variation of marker positions several cM away from their optimal positions, does not reduce much the efficacy of selection (RGC%), in particular when several markers per chromosome are used. In fact, what is important is to have at least two or three markers per chromosome, and that no chromosome is unmarked (zero marker). Given that this condition is fulfilled, the second conclusion is that selection on markers is quite efficient. In general, three or four generations of MAS are sufficient to increase RGC on non-carrier chromosomes above 99%. Hence, the gain due to selection is of about two BC generations (RGC in BC_4 with selection is approximately the same as RGC in BC_6 with no selection on markers). This gain can be economically very valuable, for example with respect to the time necessary to release new products onto the market.

Another important conclusion is that background selection is more efficient in late BC generations than in early BC generations. For example, if a BC breeding scheme is conducted over three successive BC generations, but it is preferable to genotype individuals for molecular markers at only one generation, then it is more efficient to genotype and select the individuals only in the BC_3 generation, rather than only in the BC_1 generation. This was demonstrated analytically in Hospital *et al.* (1992), and recently observed in simulations by Ribaut *et al.* (2002a). This conclusion may seem counter-intuitive, because recipient genome content is lower in BC_1, so there is 'more to select'. However, this may be explained as follows. Suppose the series of recombination events that will take place during the breeding scheme were already drawn before the start of the programme, but remained unknown to the breeder. In the BC_1 population, many chromosomal segments of donor origin are segregating. However, during the following backcross process, some of these segments will return to recipient type simply by chance. Hence, the experimental efforts devoted to the genotyping of these very segments is useless. Conversely, it is now clear that genotyping in the last generation the donor chromosome segments that were not previously removed by chance is more efficient. Hence, genotyping only the last generation could be a way to reduce the cost of the experiment. However, the efficiency of such a selection strategy will always remain below the efficiency of a strategy involving selection at every BC generation. In practice recombination events occur at random in an unpredictable manner, so that not all the genotyping efforts in early BC generations is useless, and the small gain provided can only increase the final efficiency of the breeding scheme.

2.2.2.5 Example of efficiency for a complete scheme

In order to synthesize the above conclusions, the results are presented herein of what could be a typical marker-assisted backcross breeding scheme. The results are given in Table 2.5 and were obtained by simulation of the following strategy. At each BC generation, selection was in three steps.

Table 2.5 Efficiency of a typical marker-assisted backcross scheme.

Generation	Population size	Homozygosity at selected markers (%)		RGC%	No selection
		Carrier chromosome	Non-carrier chromosomes		
BC_1	70	38.4	60.6	79.0	75.0
BC_2	100	73.6	87.4	92.2	87.5
BC_3	150	93.0	98.8	98.0	93.7
BC_4	300	100.0	100.0	99.0	96.9

Results of simulations averaged over 1000 replicates. See text for details of the selection strategy. RGC% gives the efficiency of selection over the complete genome (marker and non-marker loci). No selection gives the RGC% values if no selection on markers was performed for comparison.

1. Foreground selection: selection of all individuals that are heterozygous at the target locus (assumed controlled directly here, not by distant markers).
2. Reduction of linkage drag: selection of all individuals that are homozygous for the recipient allele at two markers flanking the target locus on each side (double recombinants) or, if no double recombinant is present in the population, selection of all individuals that are homozygous for the recipient allele at either of the two flanking markers (single recombinants).
3. Background selection on non-carrier chromosomes: selection of the one individual that is most homozygous for the most markers on non-carrier chromosomes.

Target-marker distance on the carrier chromosome was 2 cM. Each non-carrier chromosome was controlled by three markers located at optimal positions given in Table 2.4. A breeding scheme was considered which involved four BC generations, and population sizes at each generation were taken from optimal values in Table 2.3.

The results in Table 2.5 confirm that marker-assisted backcrossing is expected to be quite efficient, providing a RGC of 99% in BC_4. Again, this represents a gain of two BC generations, because a RGC of 99% would be obtained only in BC_6 with no selection on markers. It is seen from Table 2.5 that all selected markers have returned to fully homozygous recipient type in BC_4. Hence, selection on these markers would not be efficient in additional BC generations. RGC in BC_3 is already high (98%), but reduction of linkage drag is not complete at this stage, because the scheme of Table 2.5 was optimized for a total of four BC generations. If maximal efficiency in only three generations was sought, then larger population sizes should be used (see Table 2.3), providing a RGC in BC_3 of 98.5% (simulations not shown).

Note also that in practice the breeding strategy and the population sizes should be optimized with respect to the particular molecular technique used for molecular assay (e.g. Ribaut *et al.* 1997).

2.3 Genotype building strategies for multiple target genes

In some cases, several genes of interest may be manipulated at the same time. These may be genes or favourable alleles at QTL (all termed target genes here) in the same breeding scheme, with or without controlling simultaneously the genetic background into which those target genes are introduced. Again, use of molecular markers can make such breeding schemes more efficient in various ways. However, the underlying theories are still under development, and optimal strategies are not as well established as they are in the case of backcrossing for a single target. Hence, the general principles are reviewed here more briefly. The reader should refer to the cited references for more details.

Again, GB strategies here assume that an ideal genotype (ideotype) has been previously defined at a collection of loci. These may be known loci of major effects, or QTL, but in any case it is assumed that gene effects are well estimated, and sustainable, so that the selection is only at the molecular level and simply consists in screening the products of meioses (recombination) taking place in successive generations in order to obtain the ideotype as fast as possible (i.e. accumulate the favourable alleles at all previously defined loci). Strategies where the selection criterion is weighted as a function of the estimated effects of the genes considered are addressed in the next section.

2.3.1 Marker-based population screening

When several favourable genes are originally provided from two different parents, the simplest strategy involves production of an F_2, F_3, or (if possible) recombinant inbred lines (RIL) or doubled-haploid (DH) population. Then, the population is screened based on molecular markers for individuals homozygous at the appropriate loci. In this context, van Berloo and Stam (1998) have considered a set of identified QTL, each controlled by two flanking markers, and studied selection in RIL populations based on flanking markers to produce the best hybrid. If all the genes cannot be fixed in a single step of selection, it is necessary to cross again selected individuals with incomplete, but complementary, sets of homozygous loci (Charmet *et al.* 1999). However, such strategies are limited to small numbers of target loci, because the population size necessary to fix the target genes increases exponentially with the number of loci. For example, in a RIL population the frequency of homozygotes is ½ for one gene at one locus, and $(½)^k$ for k unlinked target loci – that is, less than 1 in 1000 individuals for ten target loci.

2.3.2 Marker-based recurrent selection

For even more loci, recurrent selection should be used; that is, a breeding scheme involving several generations of selection and random mating of the selected indi-

viduals. Hospital *et al.* (2000) studied selection on marker pairs flanking 50 QTL identified in an F_2 population. The best strategy seems to be to select at each generation a set of individuals that are complementary for their genotypes at flanking markers, such that each target is carried by at least two selected individuals. With this strategy, selection of three to five individuals among a total of 200 for 10 generations increases the frequency of favourable alleles at the 50 QTL up to 100% when markers are located exactly on the QTL, but only to 92% when the marker-QTL distance is 5 cM. In this case, the efficiency of marker-based selection is limited by the recombination taking place between the markers and the QTL. Hence, one has to accelerate the response to selection to fix favourable QTL alleles before marker-QTL linkage disequilibrium vanishes. The main limitation identified is the fact that selected individuals are mated at random: the authors suggest that pair-wise mating of individuals based on their marker genotypes might increase the efficiency of selection, but the theory in this area remains unexplored.

2.3.3 Marker-based gene pyramiding

When the target genes are originally present in multiple parents, it is possible to perform a marker-assisted gene pyramiding scheme, involving several initial crosses between the parents. For example, four genes (G1–G4), that are present in four different lines (L1–L4), can be combined into a single line in a two-step procedure. In the first step, two lines that are homozygous for two target genes (G1/G2 versus G3/G4) are developed by crossing pairs of lines (L1 × L2 versus L3 × L4), followed by selection of homozygotes among F_2, RIL or DH progeny. In the second step, such individuals are crossed to produce individuals that are homozygous for all four target genes. Selection of homozygotes can be on the basis of linked markers. An example of experimental implementation of such a strategy for the manipulation of QTL is given in Section 2.6. This process can be expanded to more than four genes by expanding the pyramid. However, this basic scheme is limited to small numbers of target loci. If target loci are numerous – and in particular if some loci are linked on the same chromosome – such schemes certainly require optimization, but again the theory in this area remains largely unexplored.

2.3.4 Marker-assisted backcrossing for several target genes

In the above marker-based strategies for multiple targets, attempts were not made to control the genetic background in which the target genes were accumulated (i.e. genomic regions outside the target loci). However, some traits that are improved through introgression by backcrossing might have an oligo- or polygenic basis. This is, for example, the case for the accumulation of resistance functions, or for the introgression of complex traits, because, in general, several QTL of small or medium effects account for the variability of these traits. In such cases, it is important to

control several targets *and* the genetic background in marker-assisted backcross breeding schemes.

Again, the number of individuals that must be genotyped increases exponentially with the number of target loci. Hospital and Charcosset (1997) computed such population sizes for different numbers of target QTL, and concluded that in general it is futile to plan to manipulate more than three or four QTL simultaneously in a marker-assisted backcross programme. If the targets are known loci controlled directly – and not QTL controlled indirectly by markers – then the maximum number of targets could be slightly higher, but should not exceed five or six.

However, when several targets are controlled simultaneously in one single backcross scheme, the number of individuals heterozygous at all targets is low in each BC generation, and this leaves little opportunity to select among those individuals for the genetic background (background selection). In such cases, introgression can be combined with gene pyramiding to decrease the number of individuals required (Hospital & Charcosset 1997; Koudandé *et al.* 2000). For example, if it is planned to introgress four targets, one could first perform four parallel backcross schemes each introgressing one target into the genetic background, or two backcross schemes each introgressing two targets, and then accumulate all targets in the same background by gene pyramiding (see above). This capitalizes on a higher efficiency of background selection in the separate backcross lines (provided that at least two BC generations are performed). Hence, the final efficiency after pyramiding is higher than when the targets are controlled simultaneously in one single backcross scheme. However, the total duration of the programme is then longer, because of the additional generations of pyramiding.

2.4 Selection combining molecular and phenotypic information

When the target genes do not account for all of the variability of the selected trait(s) – as would be the case for many complex traits – the gain expected from the accumulated effects of the target genes might not warrant performing selection based solely on molecular information. In such cases, it may be desirable to control both the variability accounted for by the target genes (major genes or more frequently QTL with medium to low effects) and the 'unmarked' variability. For example, de Koning and Weller (1994), Dekkers and van Arendonk (1998) or Chakraborty *et al.* (2002) have considered the optimization of MAS for identified QTL plus a possible 'polygenic' background controlling the rest of the genetic variation not explained by the identified QTL. These analyses are restricted to one or two identified QTL of sufficiently large effect.

2.4.1 The marker-phenotype index

Other methods exist in order to capitalize on all the variability accounted for by the markers, including QTL of small effects. Here, we do not aim to estimate precisely the chromosomal locations and the effects of the QTL, but simply use markers to improve the prediction of the breeding value of each individual and select the best individuals according to this value. The method is briefly outlined here, but more details can be found in Whittaker (2001).

The bases were set by the landmark paper of Lande and Thompson (1990). For a single marker, the 'molecular score' of an individual for use in recurrent selection is obtained as the estimate of the statistical association between marker genotype and phenotype. For multiple markers, genotype effects can be summed over markers into a single molecular score. Then Lande and Thompson (1990) derived an index for selection combining molecular and phenotypic information. Considering molecular score M and phenotype P as two correlated traits, the authors used classic selection index theory to compute the coefficient b_M and b_P of the index $I = b_M M + b_P P$, optimally weighting both types of information in order to maximize the genetic gain. Lande and Thompson concluded that the method was most efficient for traits with low heritability. However, Moreau *et al.* (1998) later showed that, because low heritability also reduces the power of detection of the effects associated with markers in a finite (real) population, the greatest opportunities for MAS with this method may exist for traits with moderate rather than low heritability. The efficiency of the method has been demonstrated by simulations in several papers (e.g. Gimelfarb & Lande 1994; Hospital *et al.* 1997). In all cases, population size (that must be large enough to allow a good estimate of marker effects) appears as the most critical factor limiting the efficiency of MAS. The current debate is whether only markers with significant effects should be taken into account in the molecular score, as first proposed by Lande and Thompson (1990). Moreau *et al.* (1998) showed that increasing type I error risk could increase the efficiency of MAS, by increasing the power of detection of genes with small effects. More recently, methods have been proposed that include all marker effects in the index, regardless of their statistical significance, and provide increased selection response (Lange & Whittaker 2001; Meuwissen *et al.* 2001).

One problem with this method is that QTL effects are often over-estimated, as shown by both theory (Beavis 1994; Bost *et al.* 2001) and experimentation (Melchinger *et al.* 1998). Over-estimation of QTL effects leads to too much emphasis on molecular scores in selection relative to phenotypic data, and results in a less than optimal response to selection. Alternative statistical methods for analysis of QTL data that avoid over-estimation or reduce its impact on selection response are needed (e.g. Fernando & Grossman 1989).

Another, perhaps more critical, problem is that here molecular costs are in addition to, not in place of, phenotypic costs – contrary to GB strategies with selection based solely on genotype. However, resources allocated in each generation to molecular assay could also be allocated to enhance conventional phenotypic selection (e.g. by increasing the number of individuals tested) with more profit, because the

molecular costs are still high relative to phenotypic costs (Moreau *et al*. 2000). The economic merit of MAS could be restored by reducing the frequency of re-evaluation of marker effects (Hospital *et al*. 1997). However, further work on the optimization of such strategies is required, and it is likely that the economically optimal use of MAS necessitates a complete re-thinking of the design of breeding schemes [see for example Ribaut and Hoisington (1998) for a review of changes required for plant breeding programmes].

2.5 Selection for hybrid performance

In theory, crosses between lines that are genetically more distant are expected to show greater heterosis. Genetic distance can be measured from differences in allele frequencies at anonymous markers spread throughout the genome. Evaluation of this concept for a large number of crops (Melchinger 1999) shows that marker-based prediction of hybrid performance can be efficient if hybrids include crosses between lines that are related by pedigree or which trace back to common ancestral populations. On the other hand, prediction is not efficient for crosses between lines that are unrelated or that originated from different populations, because the associations (via linkage disequilibrium) between marker loci and QTL involved in heterosis are not the same in the different populations (Charcosset & Essioux 1994).

The limited ability to predict hybrid vigour in untested crosses has motivated the development of strategies to use knowledge of QTL effects to generate crosses that are predicted to create QTL genotypes with favourable non-additive effects. An example is the use of marker-based statistical methods to predict the performance of untested crosses from performance of parental lines in a limited number of test crosses (Bernardo 1994, 1999).

2.6 Experimental results

Few results of real MAS experiments have yet been published. Some recent results in plants are presented below in order of increasing level of complexity for the use of markers, the genes manipulated and/or the traits under control.

Using markers as simple scores to hasten the recovery of recipient genome background (background selection) in backcross introgression programmes for the transfer of a single well-identified target region (direct marker) has been nicely demonstrated by the integration of the *Bt* transgene into different maize genetic backgrounds (Ragot *et al*. 1995). This confirmed the theoretical prediction that use of markers provides a gain in time of approximately two BC generations. Although few other results on this matter have been published, it is known that the technique is now widely used, in particular by private plant breeding companies.

Other experimental reports for the manipulation of known genes with indirect (linked) markers include ‘pyramiding’ of several major resistance genes in rice,

from NILs, each carrying only one gene, into a common background (Huang *et al.* 1997; Hittalmani *et al.* 2000). In all cases, control of the target genes by indirect linked markers was successful, as later demonstrated by phenotypic assay of resistance. Huang *et al.* (1997) pyramided four genes for blight resistance into different combinations (two, three or four genes) that exhibited higher levels and/or a wider spectrum of resistance than the original parents. Moreover, some pyramided lines showed resistance to pathogen races to which all parents were susceptible. Results were also generally successful for Hittalmani *et al.* (2000), who pyramided three genes for blast resistance into different combinations. However, in this case some multiple-gene combinations did not perform any better than single-gene cases, indicating that a sound knowledge of the spectrum of gene effects is necessary prior to performing the MAS program. In any case, pyramiding multiple resistance genes is a valuable step towards the development of more durable and stable crop resistance. This can hardly be achieved without the use of marker-based selection, because epistasis and/or the masking effects of genes limit the efficiency of conventional (phenotypic) breeding methods. Use of markers not only relaxes this constraint, but also provides a better understanding of interactions between these genes.

Experimental results of MAS for the manipulation of QTL (not known major genes) are less consistent. Toojinda *et al.* (1998) introgressed two QTL for stripe rust resistance in barley, through one backcross followed by one haplo-diploidization with selection on marker genotype and phenotype, into a genetic background different from that used to map the QTL. Both QTL were confirmed, and additional QTL were detected in the new background, including some resistance alleles donated by the susceptible parent. Probably those alleles were fixed in the mapping population, but this illustrates the importance of the genetic background, both for QTL detection and MAS. Han *et al.* (1997) manipulated two QTL for a component of malting quality in six-row barley, a trait that is very difficult and costly to monitor phenotypically. They screened and selected DH lines using four different strategies:

- phenotype alone;
- marker genotype alone;
- genotype followed by phenotype in tandem selection; or
- genotype and phenotype combined in an index.

All of these were either on a single-trait, or a multiple-trait basis. Results were successful for one QTL, but not for the other QTL, for which tandem and combined selection based on both marker genotype and phenotype did not perform any better than selection on phenotype alone, probably because the location of the QTL was inaccurate. However, the authors pointed out that, despite not performing any better, tandem selection provides a valuable gain in time and efforts, compared to phenotypic selection. Lawson *et al.* (1997) introgressed four target chromosomal regions containing five QTL for pest resistance (acyl sugar accumulation) from wild tomato into cultivated tomato. Starting with the introgression lines of Eshed and Zamir (1995), each carrying one target region, they performed three backcrosses followed by one

intermating generation to obtain progenies homozygous for the resistance alleles at the five QTL. Selection was based on both marker genotype and phenotype. The introgression of the four regions was successful at the genomic level. However, the level of acyl sugar accumulation in the progenies introgressed for the five QTL was lower than expected, and in particular lower than that of the interspecific F_1 hybrid, indicating that some genetic determinants (QTL) of the accumulation were missing, either lost or not controlled in the programme. Shen *et al.* (2001) manipulated four QTL for drought resistance (root depth) in rice, a trait that is very difficult to manage phenotypically. Starting from DH lines, they produced a number of BC_3F_3 lines, each introgressed for one or two QTL at most, using only selection on marker genotype, not phenotype. They re-detected and fine mapped the QTL in the progenies. Among the four QTL, one exhibited the expected effect in the progenies, one was finally revealed as a false-positive, one segment was shown to contain in fact two QTL in repulsion phase (+/–) that reduced its expression, and one segment did not exhibit the expected effect, either because the QTL was lost in the programme, or because its effect was masked by epistatic interactions. This again highlights the problems linked to the precision of the initial QTL detection with regard to the position and effect of the QTL, and the effect of possible epistatic interactions on the expression of the QTL in the progenies. Ribaut *et al.* (2002b) introgressed five target regions containing QTL for drought tolerance (reduction of ASI) in maize. The results depended on the condition of the phenotypic assay of the progenies: under stress conditions (drought), the introgressed progenies exhibited a reduced ASI, while the introgression had no visible effect in the absence of stress. Zhu *et al.* (1999) screened DH lines of barley for the presence of several QTL for yield – a very complex composite trait – based on selection for marker genotype alone. They evaluated phenotypically the progenies in five environments, including four locations and two years. The results indicated that the positions of the QTL were confirmed as correct in the progenies. However, the effects of the QTL in the progenies were often different from the expectations with regards to magnitude and direction. Moreover, the authors detected epistatic interactions between QTL, as well as numerous G × E interactions. The authors concluded that selection for complex traits should focus on allelic combinations (based on epistatic interactions) rather than on individual QTL effects.

Experimental results of MAS for QTL of the kind described above are synthesized by an experiment performed in the present author's laboratory (Bouchez *et al.* 2002). The introgression of favourable alleles at three QTL for two traits (earliness and yield) between maize elite lines by marker-assisted backcrossing showed that use of markers as simple scores to improve background selection is efficient, even with few markers, especially on non-carrier chromosomes. Foreground selection on markers to control the three target regions without the help of phenotypic assay was also efficient. However, results of the phenotypic evaluation of introgressed progenies, as well as the re-detection of QTL among those progenies depended upon the complexity of the trait under control. For the simple trait (earliness), QTL effects in the progenies were in general accordance with those expected from the original detection in the parental lines. For the more complex trait (yield) results were in

general not as good as expected, and one high-yielding allele putatively detected from the low-yielding parent finally exhibited an effect opposite to the expectation. This indicates that the estimates of QTL positions appear more reliable than the estimates of their effects, in particular with regards to genotype by environment (G × E) interactions, which were found to be significant in the experiment.

2.7 Conclusions

The application of molecular genetics in breeding programmes is currently constrained by the precision of the effects associated with markers, and the economic value of MAS. Marker-based selection is definitely useful to manipulate chromosomal regions and rapidly design new genotypes combining favourable regions (genotype building). The clear example is MAS in backcross breeding schemes for the introgression of one or a few target genes in a given genetic background. This is probably the implementation of MAS that is the most widely used in practice, in particular by private plant breeding companies, although the corresponding results are often not published. In this case, the target gene(s) could be major genes of well known, or well estimated effects. Marker-assisted backcrossing is also particularly useful for the introgression of transgenes. Here, control of the target (foreground selection) is easy because its DNA sequence is known. Moreover, it is often easier to introgress a transgene from an already genetically modified material into a new (non-modified) line, than engineer the new line.

Classical (phenotypic) selection in plant breeding is limited by the ability to estimate genetic parameters (breeding values of the individual candidates to selection) for the traits of interest, using statistical analysis of phenotypic data (quantitative genetics). The use of marker information relaxes some of the constraints of quantitative genetics selection, and provides better estimates of breeding values, by increasing the apparent heritability of the trait. However, the alleviation may be only partial, depending on the complexity of the genetic architecture of the trait. GB experiments, or selection based solely on molecular information, appears restricted to simple traits that are governed by few genes of large effects, so that genetic markers capture most genetic variation for the trait, and provide precise and sustainable estimates of breeding values. Conversely, for complex traits that are governed by several genes of medium to low effects, possibly affected by environment, it appears necessary to have an accurate evaluation of QTL effects in varying environments before initiating a GB programme. If QTL effects are not perfectly estimated and sustainable, it might be risky to perform selection based solely on markers. In such cases, selection must be on a combination of marker and phenotypic data and hence will suffer from the same limitations as conventional breeding.

Economics is the other key determinant for the application of molecular genetics in breeding programmes. Cases where the economic value of MAS is clear include situations where molecular costs are more than offset by the savings in phenotypic evaluation. Examples are the use of markers in GB programmes and selection on

traits that are costly to evaluate, but well-characterized at the molecular level (e.g. oligogenic disease resistances). In other cases, the ability to select early offsets the extra costs associated with MAS. The benefits of being able to release new genetic material more quickly can be substantial, particularly in competitive markets. The economic merit of MAS becomes questionable and more difficult to evaluate in cases where MAS is expected to provide greater genetic gain at increased costs. This is particularly the case for selection on a combination of phenotype and molecular score (see Section 2.3).

Clearly, MAS is efficient and valuable for simple traits and/or traits for which increase of genetic gain per unit of time is of high economic return. However, the advent of MAS for the ordinary breeding of complex traits relies on a re-thinking of breeding strategies, and on the availability of both statistical and molecular techniques that would provide precise estimates of gene effects in selected populations at low cost – which is far from being the case at present.

References

Beavis, W.D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in *Proceedings 49th Annual Corn and Sorghum Research Conference*, ASTA, Washington, DC.

Bernardo, R. (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science*, **34**, 20–25.

Bernardo, R. (1999) Marker-assisted best linear unbiased prediction of single-cross performance. *Crop Science*, **39**, 1277–1282.

Bost, B., de Vienne, D., Hospital, F., Moreau, L. & Dillmann, C. (2001) Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics*, **157**, 1773–1787.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. & Charcosset, A. (2002) Marker assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics*, **162**, 1945–1959.

Chakraborty R., Moreau L. & Dekkers J. C. M. (2002). A general method to optimize selection on multiple identified trait loci. *Genetics, Selection, Evolution*, **34**, 145–170.

Charcosset, A. & Essioux, L. (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. *Theoretical and Applied Genetics*, **89**, 3336–3343.

Charmet, G., Robert, N., Perretant, M.R., *et al.* (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theoretical and Applied Genetics*, **99**, 1143–1148.

de Koning, G.J. & Weller, J.I. (1994) Efficiency of direct selection on quantitative trait loci for a two-trait breeding objective. *Theoretical and Applied Genetics*, **88**, 669–677.

Dekkers, J.C.M. & van Arendonk, J.A.M. (1998) Optimum selection for quantitative traits with information on an identified locus in outbred populations. *Genetic Research*, **71**, 257–275.

Dekkers, J.C.M. & Hospital, F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Review Genetics*, **3**, 22–32.

de Vienne, D. (ed.) (2002) *Molecular Markers in Plant Genetics and Biotechnology.* Science Publishers, Inc., Enfield, NH, USA, and Plymouth, UK.

Eshed, Y. & Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, **141**, 1147–1162.

Fernando, R.L. & Grossman, M. (1989) Marker-assisted selection using best linear unbiased prediction. *Genetics, Selection and Evolution*, **21**, 467–477.

Frisch, M., Bohn, M. & Melchinger, A.E. (1999) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. *Crop Science*, **39**, 967–975.

Gimelfarb, A. & Lande, R. (1994) Simulation of marker-assisted selection in hybrid populations. *Genetical Research*, **63**, 39–47.

Groen, A.F. & Smith, C. (1995) A stochastic simulation study on the efficiency of marker-assisted introgression in livestock. *Journal of Animal Breeding and Genetics*, **112**, 161–170.

Han, F., Romagosa, I., Ullrich, S.E., Jones, B.L., Hayes, P.M. & Wesenberg, D. (1997) Molecular marker-assisted selection for malting quality traits in barley. *Molecular Breeding*, **3**, 427–437.

Hanson, W.D. (1959) Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. *Genetics*, **44**, 833–837.

Hartl, D.L. & Clark, A.G. (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates Incorporated, Sunderland, US.

Hillel, J., Schaap, T., Haberfeld, A., *et al*. (1990) DNA fingerprint applied to gene introgression breeding programs. *Genetics*, **124**, 783–789.

Hittalmani, S., Parco, A., Mew, T.V., Zeigler, R.S. & Huang, N. (2000) Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theoretical and Applied Genetics*, **100**, 1121–1128.

Hospital, F. (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics*, **158**, 1363–1379.

Hospital, F. & Charcosset, A. (1997) Marker-assisted introgression of quantitative trait loci. *Genetics*, **147**, 1469–1485.

Hospital, F. & Decoux, G. (2002) Popmin: a program for the numerical optimization of population sizes in marker-assisted backcross programs. *Journal of Heredity*, **93**, 383–384.

Hospital, F., Chevalet, C. & Mulsant, P. (1992) Using markers in gene introgression breeding programs. *Genetics*, **132**, 1199–1210.

Hospital, F., Moreau, L., Lacoudre, F., Charcosset, A. & Gallais, A (1997) More on the efficiency of marker assisted selection. *Theoretical and Applied Genetics*, **95**, 1181–1189.

Hospital, F., Goldringer, I. & Openshaw, S. (2000) Efficient marker-based recurrent selection for multiple quantitative trait loci. *Genetical Research*, **75**, 357–368.

Huang, N., Angeles, E.R., Domingo, J., *et al*. (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theoretical and Applied Genetics*, **95**, 313–320.

Koudandé, O.D., Iraqi, F., Thomson, P.C., Teale, A.J. & van Arendonk, J.A.M. (2000) Strategies to optimize marker-assisted introgression of multiple unlinked QTL. *Mammalian Genome*, **11**, 145–150.

Lande, R. & Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, **124**, 743–756.

Lange, C. & Whittaker, J.C. (2001) On prediction of genetic values in marker-assisted selection. *Genetics*, **159**, 1375–1381.

Lawson, D.M., Lunde, C.F. & Mutschler, M.A. (1997) Marker-assisted transfer of acylsugar-mediated pest resistance from the wild tomato, *Lycopersicon pennellii*, to the cultivated tomato, *Lycopersicon esculentum*. *Molecular Breeding*, **3**, 307–317.

Melchinger, A.E. (1990) Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breeding*, **104**, 1–19.

Melchinger, A.E. (1999) Genetic diversity and heterosis. In: *The Genetics and Exploitation of Heterosis in Crops* (eds J.G. Coors & S. Pandey), pp. 99–118. Crop Science Society of America.

Melchinger, A.E., Utz, H.F. & Schön, C.C. (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, **149**, 383–403.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Mohan, M., Nair, S., Bhagwat, A., *et al*. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, **3**, 87–103.

Moreau, L., Charcosset, A., Hospital, F. & Gallais, A. (1998) Marker-assisted selection efficiency in populations of finite size. *Genetics*, **148**, 1353–1365.

Moreau, L., Lemarié, S., Charcosset, A. & Gallais, A. (2000) Economic efficiency of one cycle of marker-assisted selection. *Crop Science*, **40**, 329–337.

Naveira, H. & Barbadilla, A. (1992) The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. *Genetics*, **130**, 205–209.

Ragot, M., Biasiolli, M., Delbut, M.F., *et al*. (1995) Marker-assisted backcrossing: a practical example. In: *Techniques et Utilisations des Marqueurs Moléculaires*. (Les Colloques, no. 72), INRA, Paris, France.

Ribaut, J.M. & Hoisington, D. (1998) Marker-assisted selection: new tools and strategies. *Trends in Plant Science*, **3**, 236–239.

Ribaut, J.M., Hu XueYi, Hoisington, D. & Gonzalez-de-Leon, D. (1997) Use of STSs and SSRs as rapid and reliable preselection tools in a marker-assisted selection-backcross scheme. *Plant and Molecular Biology Reports*, **15**, 154–162.

Ribaut, J.M., Jiang, C. & Hoisington, D. (2002a) Simulation experiments on efficiencies of gene introgression by backcrossing. *Crop Science*, **42**, 557–565.

Ribaut, J.M., Banziger, M., Betran, J., *et al*. (2002b) Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize. In: *Quantitative Genetics, Genomics, and Plant Breeding* (ed. M.S. Kang), pp. 85–89. CABI Publishing, Wallingford, UK.

Servin, B. & Hospital, F. (2002) Optimal positioning of markers to control genetic background in marker assisted backcrossing. *Journal of Heredity*, **93**, 214–217.

Servin, B., Dillmann, C., Decoux, G. & Hospital, F. (2002) MDM: a program to compute fully informative genotype frequencies in complex breeding schemes. *Journal of Heredity*, **93**, 227–228.

Shen, L., Courtois, B., McNally, K.L., Robin, S. & Li, Z. (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. *Theoretical and Applied Genetics*, **103**, 75–83.

Stam, P. & Zeven, A.C. (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica*, **30**, 227–238.

Tanksley, S.D. (1983) Molecular markers in plant breeding. *Plant and Molecular Biology Reports*, **1**, 3–8.

Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. (1989) RFLP mapping in plant breeding: new tools for an old science. *Biotechnology*, **7**, 257–264.

Toojinda, T., Baird, E., Booth, A., *et al*. (1998) Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theoretical and Applied Genetics*, **96**, 123–131.

van Berloo, R. & Stam, P. (1998) Marker-assisted selection in autogamous RIL populations: a simulation study. *Theoretical and Applied Genetics*, **96**, 147–154.

Visscher, P.M. (1996) Proportion of the variation in genomic composition in backcrossing programs explained by genetic markers. *Journal of Heredity*, **87**, 136–138.

Visscher, P.M., Haley, C.S. & Thompson, R. (1996) Marker-assisted introgression in backcross breeding programs. *Genetics*, **144**, 1923–1932.

Westman, A.L. & Kresovich, S. (1997) Use of molecular marker techniques for description of plant genetic variation. In: *Biotechnology and Plant Genetic Resources: Conservation and Use* (eds J.A. Callow, B.V. Ford-Lloyd & H.J. Newbury), pp. 9–48. CAB International, Wallingford, UK.

Whittaker, J.C. (2001) Marker assisted selection and introgression. In: *Handbook of Statistical Genetics* (eds D.J. Balding, M. Bishop & C. Cannings), pp. 673–695. John Wiley, New York.

Young, N.D. & Tanksley, S.D. (1989a) RFLP analysis of the size of chromosomal segments retained around the tm-2 locus of tomato during backcross breeding. *Theoretical and Applied Genetics*, **77**, 353–359.

Young, N.D. & Tanksley, S.D. (1989b) Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theoretical and Applied Genetics*, **77**, 95–101.

Zhu, H., Briceno, G., Dovel, R., *et al.* (1999) Molecular breeding for grain yield in barley: an evaluation of QTL effects in a spring barley cross. *Theoretical and Applied Genetics*, **98**, 772–779.

3 Genomic colinearity and its application in crop plant improvement

H. John Newbury and Andy H. Paterson

3.1 Introduction

Over the past 15 years it has become apparent that the organization of sets of orthologous genes within related genomes has been conserved. This phenomenon of conservation of gene order is often referred to as synteny. However, as Eckardt (2001) has pointed out, synteny is derived from the Greek words *syn* (together with) and *taenia* (ribbon) and, strictly speaking, only means that the loci are contained within the same chromosome. In spite of this, the term is often used as a synonym for colinearity which refers to conservation of gene content, order and orientation between chromosomes of different species or between non-homeologous chromosomes within a single species. Colinearity has now been clearly established in animals (e.g. O'Brien & Stanyon 1999; Dehal *et al.* 2001), in some cases over very long time periods (Mural *et al.* 2002; Smith *et al.* 2002), and evidence has been presented for its existence in prokaryotes (Horimoto *et al.* 2001). This finding that gene arrangements along the chromosomes of diverse biota is constrained by evolutionary history, together with the observation that only a tiny fraction of the theoretically possible number of gene sequences are thought to occur in nature (e.g. Dorit *et al.* 1990) comprise the foundations for 'comparative genomics,' a sub-discipline that uses the approaches of comparative biology to study genes and genomes. The exploitation of colinearity in plant genomes is mentioned in many of the chapters of this book. In this chapter, we will briefly review the history of plant colinearity studies, examine the groups of crop plants in which colinearity has been extensively studied, discuss the factors that prevent the maintenance of perfect genomic colinearity across plant taxa, and review some of the ways in which colinearity can be exploited by those involved in crop improvement programmes.

3.2 A historical perspective

Genetic mapping was given a tremendous boost in the 1980s with the advent of molecular markers, and in particular the hybridization-based restriction fragment length polymorphism (RFLP) technology. The discovery that some of the probes derived from species A could be used for RFLP analysis and genetic mapping in the closely related species B had the clear direct benefit of providing marker sets for little effort. However, this exploitation of transferable markers also revealed that the order

of marker loci in a linkage group tended to be conserved in related species. Within the Solanaceae, Bonierbale *et al.* (1988) discovered a high degree of conservation of gene order when carrying out RFLP analysis of potato and tomato, whilst Tanksley *et al.* (1988) reported more extensive rearrangement when this approach was applied to the more distantly related pepper (*Capsicum*) genome. Hulbert *et al.* (1990) employed maize DNA probes for exploratory genetic mapping of sorghum, showing clear (albeit not absolute) parallels in gene arrangement in the two distantly related grasses. At the same time as this work on pairs of closely related diploid species, there were also reports of conservation of gene order between chromosome sets that had undergone long periods of independent evolution and then been re-united in common nuclei as a result of formation of polyploid species. For example, in hexaploid wheat, which has arisen through the accumulation of the genomes from three closely related diploid relatives, Chao *et al.* (1989) were able to demonstrate conservation of gene order within the homeologous group 7 chromosomes. Similarly, Helentjaris *et al.* (1988) were able to demonstrate conserved gene order in parts of the genome of maize, a cryptic polyploid.

These seminal studies in the late 1980s led to an explosion of work in this area that has provided clear evidence for colinearity between related genomes. In these experiments, the conservation of locus organization was established by comparing the order of RFLP markers along linkage groups making use of probes that hybridize to homologous loci in pairs of genomes. The colinearity established using these techniques has been termed 'macrosynteny' (Eckardt 2001). Macrosynteny analysis based on genetic mapping is necessarily of coarse resolution, only providing the relative locations of 'milestone' loci to which the probes hybridize. This can be useful to the breeder as the transfer of information and tools from many other taxa enriches knowledge about a chromosomal interval containing a target gene in a crop species. However, each of these milestone loci may be separated by hundreds of genes and revealing the exact arrangement of genes and especially non-genic DNA in the intervening regions requires higher resolution techniques. During the past 5 years the increasing flow of genomic sequence information, particularly from *Arabidopsis* and rice, has fostered analysis of 'microsynteny', based on gene order comparisons and colinearity at the sequence level. Indeed, with the public availability of DNA sequence databases, which include annotation information as well as the identification of open reading frames, a remarkable (and still growing) amount of comparative genomic information can be obtained through the Internet, before experimental studies are considered or planned.

3.3 Examples of synteny between plant groups

3.3.1 The Brassicaceae

Numerous studies have been conducted on colinearity within the Brassicaceae, partly because of the commercial importance of members of the type genus (*Bras-*

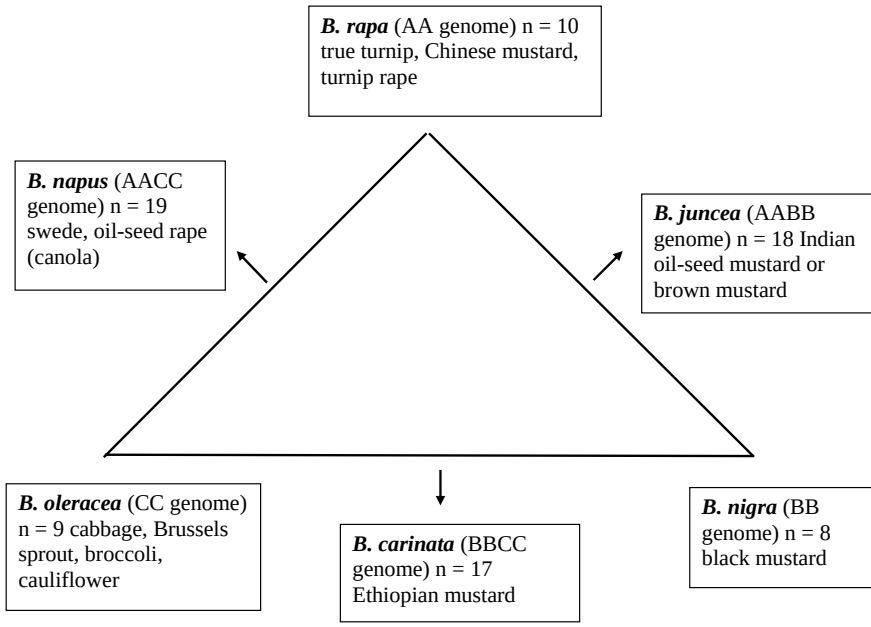


Figure 3.1 Genomic relationships within the genus *Brassica*.

sica) and more recently driven by the availability of whole genome sequence data for *Arabidopsis thaliana* which is a member of this family. Six species of *Brassica* are cultivated around the world. Three of the species are considered diploid and the genomes of these diploid species, which have differing chromosome numbers, have been designated A, B and C (Figure 3.1). There are three amphidiploid derivatives including *B. napus* (genome designation AACC) which includes oil-seed rape (canola).

A large number of genetic maps have been developed for *Brassica* covering all of the major cultivated species (reviewed by Paterson *et al.* 2000). Early genetic mapping studies revealed that a high proportion of the 'diploid' *Brassica* genomes were duplicated in *B. oleracea* (Slocum *et al.* 1990), *B. rapa* (Song *et al.* 1991) and *B. nigra* (Truco & Quiros 1994). Use of a common set of RFLP probes revealed colinear regions covering almost the entire genomes of *B. nigra*, *B. oleracea* and *B. rapa*, and virtually all available data support the notion that these 'diploid' genomes have actually been through one or more cycles of extensive chromatin duplication fairly recently in their evolutionary history. Despite this extensive macrosynteny, there have clearly been many genomic rearrangements following the differentiation of the three species (Hoenecke & Chyi 1991; Kowalski *et al.* 1994; Lagercrantz & Lydiate 1996; Truco *et al.* 1996; Lan *et al.* 2000).

Two forms of the amphidiploid *B. napus* exist. One is a 'synthetic' form deliberately produced by hybridization of the diploids *B. rapa* and *B. oleracea*, while the other is the natural wild form. The arrangement of the genome in the synthetic form is essentially unaltered from its diploid contributors (Lydiate *et al.* 1993), whilst in the natural form there have been more complicated rearrangements (Cheung & Landry 1996).

Study of the facile botanical model *A. thaliana* has allowed enormous advances in our understanding of the genetics and genomics of this member of the Brassicaceae. Building on the almost complete genome sequence of this species (e.g. Mayer *et al.* 1999; Tabata *et al.* 2000), identification of *Arabidopsis* genes is being accompanied by an increasingly valuable set of annotations proposing functions for these genes. This means that when probes (or PCR primers) derived from the *Arabidopsis* genome are used to detect loci in the genomes of other diverse plants, they are not just anonymous markers but can carry with them some functional information. The exploitation of *Arabidopsis*-derived probes within the family Brassicaceae is facilitated by a high degree of sequence homology. Sequence identity values of 92% and 93% at the nucleic acid and amino acid levels have been reported between four orthologous genes in *A. thaliana* and *Capsella rubella* (Acarkan *et al.* 2000). The average identity across 13 different *A. thaliana* and *B. napus* coding sequences at the nucleotide level coding sequences was 87% (Cavell *et al.* 1998). The degree of homology appears to vary in different gene classes since the identities reported for some *Brassica* and *Arabidopsis* disease-resistance genes (*RPM1* and *RPS2*) was higher (Grant *et al.* 1998; Wroblewski *et al.* 2000) than for a family of *CONSTANS-LIKE* genes (Lagercrantz & Axelson 2000). As expected, probes representing exon sequences are more effective in cross-hybridization studies.

In addition to the cultivated Brassicas, comparative mapping has also been useful in studying other members of the Brassicaceae. For example, the marker order found on a *Capsella* linkage group comprises two large colinear regions representing the majority of *Arabidopsis* chromosome 4 (Acarkan *et al.* 2000). *Capsella* and *Arabidopsis* are estimated to have diverged 6.2 to 9.8 million years ago. Studies have also been made of macrosynteny between the genome of *Arabidopsis* and the three *Brassica* genomes and this has been reviewed in some detail by Schmidt *et al.* (2001). *Brassica* and *Arabidopsis* are estimated to have diverged 23.1 to 25.9 (Koch *et al.* 2000) or 14.5 to 20.4 (Yang *et al.* 1999) million years ago, and there is considerable evidence for conservation of marker order between *Arabidopsis* and *B. rapa* (Teutonico & Osborn 1994), *B. nigra* (Lagercrantz 1998) and *B. oleracea* (Kowalski *et al.* 1994).

The release of the whole genome sequence of *Arabidopsis* has facilitated comparisons of Brassicaceae genomes at the DNA sequence level. Complete colinearity has been reported for homeologous regions of the genomes of *Arabidopsis* and *Capsella rubella* (Acarkan *et al.* 2000). A 60-kbp region containing at least ten different genes in *A. thaliana* was conserved with respect to gene repertoire in *C. rubella*

and for five of the genes gene order was the same in both species. Comparisons of homeologous genomic regions of *Brassica* species and *Arabidopsis* (Sadowski *et al.* 1996; Sadowski & Quiros 1998) have been complicated by the duplication (Lan *et al.* 2000) or possibly even triplication of the genome during the evolution of *Brassica*. Only if the genes of the duplicated regions are taken together is there good correlation of gene repertoires of the *Brassica* species with a corresponding region in *Arabidopsis.* In any one of the duplicated regions some of the homologues may be missing, suggesting frequent deletion events following genome duplication within *Brassica* (O'Neill & Bancroft 2000).

Perhaps surprisingly, it has been possible to detect colinearity between the genome of *Arabidopsis* and of species from which it diverged many tens of millions of years ago. Homologous blocks have been found in the genomes of sugar beet and sunflower and *Prunus* species (Barnes 2002). The sequence of a tomato bacterial artificial chromosome (BAC) clone showed conservation of gene content and order with four different segments of *Arabidopsis* chromosomes 2 to 5 (Ku *et al.* 2000). Five genes in a 57-kbp region of tomato chromosome 7 have homologous sequences in a 30-kb region of *Arabidopsis* chromosome 1, although there were two inversion events (Rossberg *et al.* 2001). Such microsynteny has also been reported between *Arabidopsis* and cotton (Paterson *et al.* 1996).

Of special interest is synteny between monocots and eudicots, the two major branches of the angiosperms. Although the monocot-dicot split has been calculated as occurring between 170 and 235 million years ago (Yang *et al.* 1999), it is also possible that limited colinearity between *Arabidopsis* and rice may be useful in the prediction of gene order (Paterson *et al.* 1996) and gene function (Devos *et al.* 1999). Early studies of comparative monocot-dicot gene order suggested the possibility of cM-sized regions over which gene order retained discernible similarity (Paterson *et al.* 1996). Mayer *et al.* (2001) showed that conservation of gene order can indeed exist over this extremely large evolutionary distance, albeit in smaller genomic regions. One-to-one comparisons (by Goff *et al.* 2002) showed discernible parallels of the partial rice genome sequence to *Arabidopsis*, although '... the conservation is less extensive than previously predicted (by Paterson *et al.* 1996)…'. However, this one-to-one comparison disregards at least one cycle of *Arabidopsis* duplication (Blanc *et al.* 2000; Paterson *et al.* 2000; Vision *et al.* 2000) plus at least one *Oryza* duplication about 40 to 50 million years ago (Goff *et al.* 2002), thereby tending to under-estimate colinearity. The finished rice sequence will permit a more definitive comparison. The very limited comparisons that have been made to date suggest that about 75% of genes in dicot species, and 40% in monocots, are found in regions that have useful colinearity with *Arabidopsis* (Barnes 2002).

3.3.2 The Fabaceae

Comparative mapping has demonstrated the existence of colinearity within many plant groups. For example, the family Fabaceae includes a very large number of crop

species many of which are grown on a large scale. These include soybean (*Glycine max*), peanut (*Arachis hypogaea*), mung bean (*Vigna radiata*), chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), common bean (*Phaseolus vulgaris*), common pea (*Pisum sativum*) and alfalfa (*Medicago sativa*). For all of these eight genera, detailed genetic maps have been constructed and evidence for colinearity between genomes is emerging. As early as 1992, Weeden *et al.* were able to establish conserved gene order across parts of the lentil and pea genomes. Evidence for the conservation of regions of the genomes within the *Vicia* genus (Menancio-Hautea *et al.* 1993; Torres *et al.* 1993) and within the *Vigna* genus (Kaga *et al.* 2000) has been presented. Boutin *et al.* (1995) revealed the existence of a high degree of colinearity between mungbean (*V. radiata*) and common bean (*P. vulgaris*) linkage maps; comparisons between these linkage groups and that of soybean showed that short sections of the soybean linkage map exhibited colinearity. Humphry *et al.* (2002) established a high level of colinearity between the linkage maps of lablab (*Lablab purpureus*) and mungbean (*V. radiata*).

All of these studies made use of molecular markers that can be transferred between species and/or genera and hence provide evidence at the macrosynteny level. Establishing the extent of microsynteny will be achieved as more sequence information becomes available for legume species. Soybean is the most heavily studied of the crop legumes, especially in relation to disease resistance genes. Lee *et al.* (2001) examined homologous regions of the genomes of soybean, common bean and mungbean and showed that there is not only conservation of large regions of the genomes but that these conserved linkage blocks are also represented twice in the soybean genome. Soybean BAC libraries are quite well developed and are being grouped into contigs and subjected to selective sequencing (Marek *et al.* 2001). The acquisition of BAC-end sequences has allowed the testing for microsynteny between the genome of soybean and other genomes of other plant species.

Because of the many diverse species within the family Fabaceae that are important as crop plants, it has been proposed that model legume species be intensively studied to provide genetic and genomic information that can inform studies on the lesser studied taxa. *Medicago truncatula* is becoming established as such a model partly because of its small genome (Cook 1999). Microsynteny between pea (*Pisum sativum*) and *M. truncatula* has been demonstrated using a fine-structure genetic map of orthologous genomic regions (Gualtieri *et al.* 2002); microsynteny has also been established between regions of the genomes of *M. truncatula* and soybean (Marek *et al.* 2001). Large-scale international projects for *M. truncatula* genomics have been initiated and essential tools are being developed for structural genomics (genome mapping, BAC libraries, genome sequencing) and functional genomics [expressed sequence tags (ESTs), microarrays, mutant collections] (Journet *et al.* 2001). *M. truncatula* promises to become a nodal taxon for gene identification in crop legumes. However, advances have also been made with another possible model legume species, *Lotus japonicus* (Ito *et al.* 2000; Kawasaki & Murakami 2000). This species has become the focus of a project that aims to sequence one-third of

its entire genome within 5 years (Colebach *et al*. 2002). Such sequence information will be invaluable in the search for agronomically important genes in less well-studied crop legumes if colinearity between genomes can be characterized. It will be important to develop good genetic maps of poorly studied legume crops along with transferable markers that allow macrosynteny to be established so that information can flow from the models into what are sometimes referred to as 'orphan' crops. Genomic information on *M. truncatula* and *L. japonicus* is also expected to be of particular interest with respect to genes involved in endomycorrhizal and nitrogen-fixing root symbioses (Stougaard 2001; Colebach *et al*. 2002) as these associations do not occur in either *Arabidopsis* or rice.

Exploratory efforts in genomic sequencing hint at legume-*Arabidopsis* microsynteny in at least some genomic regions. BAC-derived DNA sequence data show that microsynteny exists between regions of the soybean and *A. thaliana* genomes (Marek *et al*. 2001; Foster-Hartnett *et al*. 2002). Grant *et al*. (2000) showed that synteny between soybean and *Arabidopsis* existed by using conceptual translations of DNA sequences from loci that map to three soybean linkage groups. Linkage group A2 and *Arabidopsis* chromosome 1 showed significant colinearity over much of their lengths. Lee *et al*. (2001) have reported that genomic regions showing conserved gene order in three legume genomes are also relatively conserved in *Arabidopsis*.

3.3.3 The Poaceae

Any review of genomic colinearity in plants would be incomplete without mention of the studies that have been carried out among the grasses. However, since the subject has been extensively reviewed (Gale & Devos 1998; Devos & Gale 2000; Freeling 2001; Feuillet & Keller 2002) and since comparative mapping is referred to in the chapters on maize and wheat within this volume, only an outline is given in this chapter. The grasses (Poaceae), of course, contain many cereal species that played an enormous role in the transition of humans from hunters to farmers, and these remain the main providers of human food on earth. The relatively close relationship among the major cereals, comprising a branch of the plant family tree that may be as little as 70 million years old, has fostered extensive comparative genomic study.

The concept of genomic colinearity in the cereals began with the reports, mentioned earlier, of Helentjaris *et al*. (1988) on maize (a cryptic polyploid) and Chao *et al*. (1989) on hexaploid wheat. Since then, the transfer of hybridization probes between genera has provided clear evidence for macrosynteny between the genomes of quite widely separated cereal crop plants; this has been graphically displayed by the organization of cereal genomes into concentric circles so that ostensibly colinear regions in different species lie along any radius (Moore *et al*. 1995). In this representation, the rice genome has been used as the inner circle against which all of the other genomes are aligned by colinearity. This '*Oryza*-centric' view of the

cereals is justified by the relatively small size of the rice genome and the enormous advances in rice genomics that have been achieved through international efforts over the last two decades. The importance of rice as a model for comparative genomics has recently been reviewed by Shimamoto and Kyozuka (2002). Briefly, the complete rice genome has recently been 'shotgun-sequenced' by both an industrial consortium (Goff *et al.* 2002) and a public effort (Yu *et al.* 2002), and an international collaboration is working toward a truly 'finished' sequence. In addition, there is an extensive EST database. Transposon- and T-DNA-tagging have been used to produce genotypes in which single genes are mutagenized, and these can be used in functional genomics programmes. Extensive genetic mapping also allows the isolation of novel genes by map-based cloning. Microarray technology for studying mRNA expression profiling is available. The production of transgenic rice plants is now relatively easy compared with other major cereals.

The alignment of cereal genomes on the basis of colinearity started with pairwise comparisons between genetic maps (see Gale *et al.* 2001). Hulbert *et al.* (1990), in their partial exploration of maize-sorghum colinearity, were perhaps the earliest example. Later, Ahn and Tanksley (1993) were able to align maps of rice and maize, and colinearity was also found between the genomes of rice and wheat (Kurata *et al.* 1994; van Deynze *et al.* 1995a). Further studies revealed colinear relationships between commercially less significant cereals, including sugar cane (*Saccharum officinale*, Paterson *et al.* 1995; Glaszmann *et al.* 1997), sorghum (*Sorghum bicolor*, Paterson *et al.* 1995), foxtail millet (*Setaria italica*, Devos *et al.* 1998), oats (*Avena* spp., van Deynze *et al.* 1995b), pearl millet (*Pennisetum glaucum*, Devos *et al.* 2000) and wild rice *(Zizania palustris*, Kennard *et al.* 1999). It is expected that the genomes of fescue (*Festuca* spp.), rye grass (*Lolium* spp.), tef (*Eragrostis tef*) and finger millet (*Eleusine coracana*) will be added to the colinearity-based alignment diagrams very soon (Gale *et al.* 2001).

The availability of the colinearity relationships, especially with the model genome of rice, is of particular value for species that have much larger genomes (such as wheat and sugar cane) that are less amenable to genetic study. An understanding of the relationships is also of critical importance for 'orphan crops'. For example, foxtail millet has received very little attention from geneticists. The relatively simple procedure of producing a genetic map for this species using markers that allow alignment with rice (Devos *et al.* 1998) has allowed a flow of information about putative locations of genes controlling agronomic traits in rice, maize or wheat for which the physiological, genetic and biochemical control is already elucidated. The process has been facilitated by the development of various relational databases designed to provide integrative information about maps, sequence, genes, genetic markers, mutants, quantitative trait loci (QTL), controlled vocabularies and publications. A simple example of the alignment of the genomes of rice and wheat is shown in Figure 3.2.

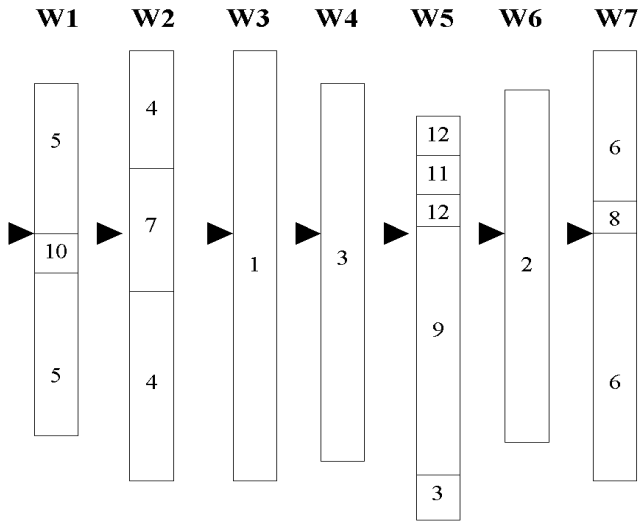


Figure 3.2 The genome of wheat (chromosome 1–7) in which the colinear regions of rice chromosomes (1–12) are marked. (After Gale & Devos 1998.)

3.4 The limits to colinearity

Even before the availability of large-scale sequencing data for model plant species, it had become apparent that plant genomes had been subjected to duplication events during their evolution. This is clear even within *Arabidopsis*, which was selected as a model plant largely on the basis of its small genome (Kowalski *et al.* 1994; AGI 2000; Blanc *et al.* 2000; Paterson *et al.* 2000; Vision *et al.* 2000). As has already been mentioned, many current 'diploid' species, such as maize and *Brassica* species, have an ancient polyploid origin. Although duplications at the whole genome level are the most obvious, some have also occurred at the single gene level, at the level of chromosomal segments or of whole chromosomes. The subsequent divergence of structure and function of individual 'partner' genes, sometimes into large gene families, along with changes in local gene order due to inversion and translocation events, complicates the process of identifying colinear genomic regions (Bennetzen 2000; Eckardt 2001). Orthologues are genes from different species that are derived from a common ancestor, whereas paralogues are genes within a species that have arisen from a duplication event. Since most genes are present as multiple homologues within a plant genome it can be very difficult to determine whether truly orthologous loci are being compared.

Even when comparisons are carried out at the sequence level (microsynteny) the identification of orthology can be confusing. Shimamoto and Kyozuka (2002) have carried out detailed structural and functional comparisons of genes involved in a range of biological processes in rice and *Arabidopsis*. They have shown that apparently orthologous genes do not always have similar functions. The numbers and

kinds of members of gene families are frequently different between the two species. Importantly, homologous genes are often differently regulated in different species.

A further problem in the exploitation of colinearity is that of gene deletion. For example, while most of the *Arabidopsis* genome appears to have been through at least one, and possibly more than one, cycle of chromatin duplication, only about 30% of the genes in modern *Arabidopsis* retain duplicate copies (Blanc *et al.* 2000; Paterson *et al.* 2000; Vision *et al.* 2000), the remainder presumably deleted as a part of a 'diploidization' process (Lynch & Conery 2000). Fine mapping has already shown the apparent deletion of self-recognition genes in *Arabidopsis* (Conner *et al.* 1998), and there is evidence of the deletion of disease-resistance genes from some species (see later). This means that the search for orthologous genes in colinear regions of genomes will sometimes be fruitless.

3.5 How can synteny be exploited by those attempting to improve plants?

Before discussing the direct applications of colinearity to crop plant improvement, it is worth considering some of the indirect ways in which genomic relationships can inform molecular genetic studies. For instance, the knowledge that regions of two genomes are colinear can be valuable in the process of sequence annotation. Where two species diverged over 5 million years ago, only the genes are expected to be extensively conserved. Hence, sequences shared between two orthologous regions are likely to be genes (Avramova *et al.* 1996; Tikhonov *et al.* 1999). The demonstration that two genes are orthologous can also be extremely useful in defining intron/exon boundaries. Where these are known in one species, information can often be transferred to another less well-studied species. Colinearity studies provide valuable information about the processes of genome evolution (e.g. Heslop-Harrison 2000; Schmidt 2002) and may help to determine the nature of the evolved functions that make one species different from another (Bennetzen and Freeling 1997).

There are several direct benefits of knowledge of colinearity relationships between plant species for those involved with crop improvement. At its crudest, it allows the prediction of the location of genes controlling a particular function in a crop species. Information about colinear regions of genomes may improve prospects of the transfer of such traits via homeologous recombination in wide crosses (e.g. Jaffe *et al.* 2000). It may also provide markers that can be transferred from a well-mapped to a less well-studied species. The availability of these markers may allow saturation mapping of a target genomic region of a crop species and facilitate both marker-assisted selection in breeding programmes or map-based cloning. Alternatively, colinearity information at the sequence level may allow the cloning and characterization of an orthologue of a crop gene from another species, such as rice or *Arabidopsis*, for which physical mapping is more advanced. Having cloned a locus of agronomic importance, colinearity also provides the possibility of accumulating alleles of that locus from more and more distantly related species. These can be used

to inform studies on protein function (e.g. identification of conserved regions which are presumed to have important functions) and serve as a catalogue of genes that, in the future, could be transferred by genetic engineering in order to fine-tune aspects of performance. Specific examples of these applications of comparative mapping will be encountered during the consideration of three agronomic traits.

3.5.1 Disease resistance

Since the first map-based cloning of a plant disease resistance gene (Martin *et al.* 1993), dozens of *R* genes have been isolated (Baker *et al.* 1997) along with many more *R*-like genes for which specific functions have not been ascribed (Leister *et al.* 1998; Cannon *et al.* 2002). Specific genes allow the plant to recognize, and mount defence reactions against, some genotypes of species of either fungi, bacteria, nematodes or viruses. The majority of *R* genes encode a protein with a nucleotide-binding site (NBS) that occupies a position N-terminal to a leucine-rich repeat LRR (van der Beizen & Jones 1998). The NBS proteins can be grouped into two classes. The first class includes those with the toll/interleukin-1 receptor homology N-terminal to the NBS (the TIR class) and includes resistance genes *N* from tobacco, *M* and *L6* from flax and *RPP5* from *Arabidopsis.* The second class are those without the toll/interleukin-1 receptor homology (the non-TIR class) and include genes *RPS2* and *RPM1* from *Arabidopsis*, *I2*, *MI* and *PRF* from tomato and *DM3* from lettuce (Michelmore 2000). The conservation of function of parts of the *R* gene products has allowed the development of sets of redundant primers that allow the amplification of putative *R* genes using PCR. Further large sets of putative *R* genes have been identified during whole genome sequencing programmes. NBS sequences occur in large numbers in the *Arabidopsis* genome and tend to be clustered (Botell *et al.* 1997). Many phenotypically defined resistance genes map to clusters on chromosomes IV and V (Kunkel 1996).

A central problem for those seeking to improve disease resistance is the determination of the specificities of the large numbers of *R* genes present in crop plant genomes so that appropriate sequences can be deployed effectively. This requires careful experimentation involving a combination of mutation and transgenic analyses (Michelmore 2000), and care is needed since ectopic expression of resistance genes may result in non-specific resistance, as has been found for *PRF* or *PTO* (Oldroyd & Staskawicz 1998; Tang *et al.* 1999). It is possible that comparative genomics will eventually provide a series of allelic *R* gene variants that can be exploited in plant improvement programmes. However, so far only a few domain swap experiments have been reported (Ellis *et al.* 1999), and the possibility of deliberately creating a new resistance gene specificity remains far in the future.

Meanwhile, much of the work in this area has relied on conserved regions of protein sequence for the identification of *R* gene candidates. There have been some attempts to use genomic colinearity to help find orthologous *R* genes, but problems have been encountered. For example, Leister *et al.* (1998) used conserved domains within dicot NBS-LRR *R* genes to amplify related sequences from monocots using

PCR. Mapping of these sequences in rice and barley showed linkage to characterized *R* genes. However, interspecific analyses of *R*-like genes frequently revealed non-syntenic map locations between the cereals rice, barley and foxtail millet. The data suggest a dramatic rearrangement of *R* gene loci between related species that has not been encountered with other classes of genes. Studies within the Brassicaceae also provide cause for concern with regard to the usefulness of colinearity for the identification of orthologous *R* genes using colinearity. The *R* gene *RPM1* confers resistance to certain races of *Pseudomonas syringae* pv. *maculicola* in *Arabidopsis*. Grant *et al.* (1995) showed that the difference between resistant and susceptible lines of *Arabidopsis* arose from an intraspecific insertion or deletion of the *RPM1* gene. They later demonstrated that whilst colinearity of genes flanking *RPM1* is conserved between the genomes of *Arabidopsis* and *Brassica napus*, in some cases the *RPM1* locus is absent (Grant *et al.* 1998).

Disease resistance genes appear to defy the general colinearity seen between chromosomes in grass species; NBS-LRR sequences show copy number variation of each gene class within a species and limited conservation of sequence and position between species (Leister *et al.* 1998; Michelmore 2000). It appears that high rates of divergence occur for these genes and they may be eliminated at a high frequency.

Attempts to use colinearity with rice as part of a map-based cloning strategy for the barley *RPG1* gene suffered from a similar problem and serve as a good example of the difficulties that can arise in such projects. Stem rust caused by *Puccinia graminis* f. sp. *tritici* was among the most devastating diseases of barley in the northern Great Plains of the USA and Canada before the deployment of the *RPG1* gene. Since then it has provided durable resistance in a wide range of barley cultivars. Initial studies demonstrated a high degree of microsynteny between the telomeric regions of barley chromosome 1P (in which *RPG1* maps) and rice chromosome 6. An immediate benefit was the exploitation of probes that had previously been used for mapping that region in rice to be mapped in barley (Kilian *et al.* 1995). This allowed a high-resolution map of the barley *RPG1* region to be established, with 1400 gametes yielding a map density of 3.6 markers per 0.1 cM (Kilian *et al.* 1997). In an attempt to identify a rice orthologue for *RPG1*, a 35-kb region of a rice BAC clone was sequenced. However, a rice *RPG1* homologue was not present in this colinear region (Han *et al.* 1999). Cloning of the *RPG1* gene was eventually achieved by map-based cloning in barley (Brueggeman *et al.* 2002). Thus, whilst colinearity did not allow the cloning of an orthologous gene in model species (because the gene appears to be absent from the rice genome), it did allow the transfer of large numbers of markers to allow production of the high-resolution barley genetic map that allowed gene isolation.

This colinearity-informed provision of molecular markers has also been reported in other *R* gene studies. For example, Jaffe *et al.* (2000) proposed that colinearity between *Hordeum bulbosum* and barley (*H. vulgare*) gives a basis for the development of marker-mediated introgression through intraspecific hybridization between the two species. Asnaghi *et al.* (2000) identified a major rust-resistance gene in a sugarcane cultivar that was linked to a specific marker. Sugarcane cultivars have a very complex genome with about 10 to 12 homologous or homeologous versions of

each chromosome. Exploitation of colinearity between the genomes of sugarcane, sorghum, maize and rice led to the identification of homeologous chromosome segments in these three relatives. This allowed the addition of 19 RFLP markers shown to be linked to the resistance gene on the sugarcane genetic map. These studies indicate that it is possible to transfer mapping resources from well-characterized genomes to less well-studied and/or complex genomes in order to facilitate marker-assisted selection or to saturate regions of the genetic map with a view to map-based cloning.

3.5.2 Flowering time

The timing of flowering – in some crops referred to as heading date – is a key target trait in breeding programmes for many crop species. It is clear that multiple environmental and endogenous inputs regulate flowering, and studies with *Arabidopsis* have allowed the identification of an integrated network of pathways that control the switch from a vegetative to a reproductive mode of development (Simpson & Dean 2002). Much effort has been expended in identifying and characterizing individual *Arabidopsis* genes that influence flowering time, and homologues of these genes have been found in other species. Because of their close phylogenetic relationship, information about *Arabidopsis* flowering time genes can be readily exploited in *Brassica* species. It has been proposed that different strategies can be used to exploit the close genomic relationships and, in particular, to correlate *Brassica* QTL with *Arabidopsis* genes (Schmidt *et al*. 2001). First, putative candidate *Arabidopsis* genes can be used as markers on *Brassica* mapping populations to see whether the segregation of the RFLP alleles coincides with that of QTL for trait performance. Alternatively, *Brassica* markers near a particular QTL can be used to find colinear regions of the *Arabidopsis* genome to suggest candidate genes.

The *Arabidopsis* gene *CONSTANS*, mutants of which show delayed flowering only under long-day conditions, has been cloned and shown to encode a protein with the zinc finger motifs associated with some transcription factors (see Coupland 1995). It has been shown that homologues of *CONSTANS* are co-inherited with QTL for flowering time in *Brassica nigra* (Lagercrantz *et al*. 1996). Eleven genomic sequences derived from a 1.5-Mbp region of the *Arabidopsis* genome around the *CONSTANS* locus were used as RFLP probes on a mapping population of *B. nigra* genotypes and identified two colinear regions. Since one of the 11 RFLP probes used was the *CONSTANS* locus itself, the identification of its homologues in *B. nigra* primarily made use of sequence homology. However, the finding that there is striking fine-scale colinearity between duplicated regions of the *Brassica* genome and the region of the *Arabidopsis* genome carrying the *CONSTANS* locus served to strongly confirm the identification of the two orthologous loci.

The *B. nigra CONSTANS* homologue on linkage group 2 was coincident with the most likely position of the major QTL controlling flowering time in this mapping population (the single locus accounting for 53% of the total variation in this trait). The results also suggested that a second *CONSTANS* homologue, on linkage group

8, might account for 12% of the variation in flowering time. This implication of *CONSTANS* as a determinant of flowering time in *Brassica* has been extended by a study by Bohoun *et al.* (1998). They showed that the region of *Arabidopsis* chromosome 5 containing *CONSTANS* is colinear not only with duplicated regions of the *B. nigra* genome, but also with sections of linkage groups in *Brassica oleracea* that have been shown to contain flowering time QTL (Bohoun *et al.* 1998). As with all QTL analyses, one must bear in mind that the QTL detected depend on the presence of allelic differences at potentially important loci between the parents of the mapping population and also on the environmental conditions used. For *CONSTANS*, the environment applied is critically important since mutations at this locus alter flowering time under long days but have no effect under short-day conditions.

Other studies have shown colinearity between genomic regions containing two flowering time QTL in *B. napus* and *B. rapa* (Osborn *et al.* 1997). Strong colinearity between parts of these genomes is to be expected, since *B. rapa* is hypothesized to be one of the diploid parents of the amphidiploid *B. napus.* The *B. rapa* and *B. napus* regions containing the first pair of flowering time QTL was colinear with the top of *Arabidopsis* chromosome 5 that contained the flowering time genes *CONSTAN*S, *FLC* and *FY.* The backcrossing of the 'late' *B. rapa* allele into an early flowering line resulted in monogenic segregation of flowering time in a segregating population. DNA probes that were shown previously to detect RFLP loci in the appropriate region of the *B. rapa* genome, or in the colinear regions of *B. napus* or *A. thaliana* (including two clones representing flowering time genes), were used to construct a high-resolution map around the *B. rapa* QTL. An RFLP detected by an *A. thaliana* cDNA clone of *FLC* (flowering locus C) exactly co-segregated with the QTL for the 414 gametes analysed, strongly suggesting that this *B. rapa* QTL is a *FLC* homologue (Kole *et al.* 2001). Michaels and Amasino (1999) proposed that *FLC* acts through a rheostat-like mechanism to control flowering time, and that modulation of *FLC* expression is a component of the vernalization response. The *B. rapa* and *B. napus* regions containing the second pair of flowering time QTL showed fractured colinearity with several regions of the *Arabidopsis* genome, including the top of chromosome 4 where the flowering time gene *FRIGIDA (FRI)* is located. Dominant alleles of *FRI* confer late flowering, which is reversed to earliness by vernalization (Johanson *et al.* 2000).

These studies indicate that colinearity can play an important role in dissecting the genetic control of flowering time within the *Brassica* genus by using *Arabidopsis* genetic data. However, colinearity has also been exploited within the cereals for the identification of flowering time loci. The *PPD-H1* photoperiod response gene plays a major role in flowering time in barley, and its genomic location suggests that it is homeologous to the wheat *PPD* gene series. Dunford *et al.* (2002) attempted to use colinearity relationships to help isolate this barley gene. A region of the barley chromosome 2HS bears the *PPD-H1* gene, and this is colinear with part of rice chromosome 7L which bears the *HD2* heading date gene. Their comparative mapping analysis provided valuable additional markers for the *PPD-H1* region but, because of the remnants of an ancient duplication that had apparently occurred before the

divergence of these species, they were unable to identify short chromosome intervals bearing candidate genes for further investigation.

Snape *et al.* (2001) showed that genes controlling vernalization response and cold tolerance in wheat mapped to chromosomes 5A and 5D. A syntenous region of the rice genome was identified, and it was proposed that this region should be identifiable in all grasses. This suggests that the tagging of such syntenous regions with molecular markers will enable plant breeders to carry out marker-assisted selection for alleles conferring improved performance for these traits.

3.5.3 Plant height

During the 1960s and 1970s, cereal breeders substantially increased grain yield partly by breeding shorter varieties in what has been called the 'Green Revolution'. Mutant alleles at specific loci in wheat and rice were introduced into new varieties that played an important role in feeding increased human populations. In recent years, some of the genes that were exploited by breeders have been characterized and, with the availability of colinearity and sequence homology information, a range of alleles at these loci have been identified and can be utilized in future plant improvement programmes.

Many of the 'dwarfing' genes that have been identified encode products that play a role in gibberellin biology – either in GA synthesis or GA perception. The *A. thaliana* genome contains the gene *GAI* (gibberellin insensitive), the wild-type allele of which encodes a protein containing features expected of a transcription factor. A mutant allele encodes a mutant protein lacking 17 amino acids from near the amino terminus and the phenotype of the mutant shows reduced plant height, reduced response to added gibberellin and an increase in *in planta* gibberellin levels. These phenotypic characteristics are shared by mutants at the *RHT-B1b* and *RHT-D1b* loci (*RHT* = reduced height) in wheat and at the *D8* locus in maize. Peng *et al.* (1999) used database searches of rice ESTs to find a clone encoding an amino acid sequence nearly identical to part of the *Arabidopsis* wild-type protein. This EST was used to isolate a homologous wheat cDNA that hybridized to homeologous regions of the 4A, B and D chromosomes suspected to contain the *RHT-1* locus. Comparative mapping demonstrated that the wheat *RHT-1* region is colinear with the maize region containing *D8* and a region of the rice genome. The amino acid sequences of *RHT-1*, *D8* and *GAI* and *RGA* (which encodes another *Arabidopsis* gibberellin signalling protein) showed considerable homology. Ivanic *et al.* (1999) have independently provided strong evidence for the homologous nature of the dwarfing barley gene *DWF2* which is located in a genomic region colinear with the wheat *RHT-B1* and *RHT-D1* loci.

These studies, which have exploited both genomic synteny and gene (protein) sequence homology, have provided an allelic series of semi-dominant mutations that confer different severities of dwarfism. It is proposed that dwarfing alleles of *GAI*, *RHT-1*, *D8* or *DWF2* can be used directly to adjust plant height in crop plants.

As proof of principle, *GAI* constructs were introduced into Basmati 370 rice, which is a tall variety that is highly susceptible to damage by wind and rain (Peng *et al.* 1999). Previous attempts to reduce the height of this variety had not been successful since it was not possible to retain its unique culinary qualities in a shorter genotype. However, this was achieved by genetic engineering with a *GAI* allele. This is a clear example of the exploitation of colinearity (and gene sequence homology) to provide a range of alleles that can be exploited by transgenic methods.

It should be noted that the parallel 'Green Revolution' breeding breakthrough in rice also involved exploitation of a mutant allele of a gene involved in gibberellin biology (*SD1*), and the function of this gene has been fully characterized both genetically and biochemically (Sasaki *et al.* 2002). In this case, the gene product is involved in a step in gibberellin synthesis rather than gibberellin perception (so that the mutant is gibberellin-sensitive). Syntenic relationships between plant stature genes have also been demonstrated with the dominant GA-sensitive dwarfing genes *RHT12* in wheat and *DDW1* in rye (Korzun *et al.* 1997; Børner *et al.* 1998). The further elucidation of allelic series from homeologous 'plant stature' loci derived from an increasing range of species should allow the fine-tuning of plant height of any plant species that can be genetically engineered.

3.6 Conclusion

There is now overwhelming evidence that the organization of sets of orthologous genes within related genomes has been conserved within angiosperms. The level of conservation of gene order is highest between taxa that diverged most recently (i.e. those that are more closely phylogenetically related). Difficulties in aligning orthologous genomic regions occur for several reasons. Duplication events can occur at the level of the whole genome (in polyploids) or for chromosome segments. Along with subsequent inversion and translocation events, this can complicate the process of identifying colinear genomic regions. The situation is made even more difficult because of duplications or deletions of individual genes in one of the phylogenetic lines being studied. In spite of these problems, colinearity between plant genomes can provide valuable information to those involved in plant improvement, especially by allowing information obtained in heavily studied species such as *Arabidopsis* and rice to flow to understudied crop species. It can suggest chromosomal regions that may carry genes of agronomic importance. It can provide sets of DNA probes from these regions of heavily studied genomes that can be used to develop polymorphic markers in colinear regions of less well-studied genomes. These markers can then be exploited for fine genetic mapping of a target gene (including QTL) and marker-assisted selection during breeding programmes. Alternatively, the markers can facilitate map-based cloning and the accumulation of a series of alleles for agronomically important genes that can potentially be transferred using transgenic technology.

References

Acarkan, A., Rossberg, M., Koch, M. & Schmidt, R. (2000) Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant Journal*, **23**, 55–62.

AGI (The Arabidopsis Genome Initiative) (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Ahn, S.N. & Tanksley, S.D. (1993) Comparative linkage maps of the rice and maize genomes. *Proceedings of the National Academy of Sciences, USA*, **90**, 7980–7984.

Asnaghi, C., Paulet, F., Kaye, C., *et al.* (2000) Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics*, **101**, 962–969.

Avramova, Z., Tikhonov, A., SanMiguel, P., *et al.* (1996) Gene identification in a complex chromosomal continuum by local genomic cross referencing. *Plant Journal*, **10**, 1163–1168.

Baker, B., Zambryski, P., Staskawicz, B. & Dinesh-Kumar, S.P. (1997) Signaling in plant-microbe interactions. *Science*, **276**, 726–733.

Barnes, S. (2002) Comparing *Arabidopsis* to other flowering plants. *Current Opinion in Plant Biology*, **5**, 128–133.

Bennetzen, J.L. (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell*, **12**, 1021–1029.

Bennetzen, J.L. & Freeling, M. (1997) The unified grass genome: synergy in synteny. *Genome Research*, **7**, 301–307.

Blanc, G., Barakat, A., Guyot, R., Cooke, R. & Delseny, M. (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell*, **12**, 1093–1101.

Bohoun, E.J.R., Ramsay, L.D., Craft, J.A., *et al.* (1998) The association of flowering time quantitative trait loci with duplicated regions and candidate loci in *Brassica oleracea*. *Genetics*, **150**, 393–401.

Bonierbale, M.D., Plaisted, R.L. & Tanksley, S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosome evolution in potato and tomato. *Genetics*, **120**, 1095–1103.

Børner, A., Korzun, V. & Worland, A.J. (1998) Comparative genetic mapping of loci affecting plant height and development in cereals. *Euphytica*, **100**, 245–248.

Botell, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T., Jones, J.D.G. & Somerville, S.C. (1997) Map position of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. *Plant Journal*, **12**, 1197–1211.

Boutin, S.R., Young, N.D., Olson, T.C., Yu, Z.-H., Shoemaker, R.C. & Vallejos, C.E. (1995) Genome conservation among three legume genera detected with DNA markers. *Genome*, **38**, 928–937.

Brueggeman, R., Rostoks, N., Kudrna, D., *et al.* (2002) The barley rust resistance gene *RPG1* is a novel disease resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences, USA*, **99**, 9328–9333.

Cannon, S.B., Zhu, H.Y., Baumgarten, A.M., *et al.* (2002) Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *Journal of Molecular Evolution*, **54**, 548–562.

Cavell, A.C., Lydiate, D.J., Parkin, I.A.P., Dean, C. & Trick, M. (1998) Collinearity between a 30-centiMorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica* genome. *Genome*, **41**, 62–69.

Chao, S., Sharp, P.J., Worland, A.J., Warham, E.J., Koebner, R.M.D. & Gale, M.D. (1989) RFLP-based wheat homeologous group 7 chromosomes. *Theoretical and Applied Genetics*, **78**, 495–504.

Cheung, W.Y. & Landry, B.S. (1996) Current status of genome mapping in the Cruciferae. In: *Genome Mapping in Plants* (ed. A.H. Paterson), pp. 193–210. Landes Bioscience Press, Austin, USA.

Colebach, G., Trevaskis, B. & Udvardi, M. (2002) Symbiotic nitrogen fixation research in the postgenomics era. *New Phytologist*, **153**, 37–42.

Conner, J.A., Conner, P., Nasrallah, M.E. & Nasrallah, J.B. (1998) Comparative mapping of the *Brassica S* locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. *Plant Cell*, **10**, 801–812.

Cook, D.R. (1999) *Medicago truncatula* – a model in the making! *Current Opinion in Plant Biology*, **2**, 301–304.

Coupland, G. (1995) Genetics and environmental control of flowering time in *Arabidopsis*. *Trends in Genetics*, **11**, 393–397.

Dehal, P., Predki, P., Olsen, A.S., *et al.* (2001) Human chromosome 19 and related regions in mouse: Conservative and lineage-specific evolution. *Science*, **293**, 104–111.

Devos, K.D. & Gale, M.D. (2000) Genome relationships: the grass model in current research. *Plant Cell*, **12**, 637–646.

Devos, K.M., Wang, Z.M., Beales, J., Sasaki, T. & Gale, M.D. (1998) Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theoretical and Applied Genetics*, **96**, 63–68.

Devos, K.M., Beales, J., Nagamura, Y. & Sasaki, T. (1999) *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot monocot divide? *Genome Research*, **9**, 825–829.

Devos, K.M., Pittaway, T.S., Reynolds, A. & Gale, M.D. (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theoretical and Applied Genetics*, **100**, 190–198.

Dorit, R.L., Schoenbach, L. & Gilbert, W. (1990) How big is the universe of exons? *Science*, **250**, 1377–1382.

Dunford, R.P., Yano, M., Kurata, N., *et al.* (2002) Comparative mapping of the barley *PPD-H1* photoperiod response gene region, which lies close to a junction between two rice linkage segments. *Genetics*, **161**, 825–834.

Eckardt, N.A. (2001) Everything in its place: conservation of gene order among distantly related plant species. *Plant Cell*, **13**, 723–725.

Ellis, J.G., Lawrence, G.J., Luck, J.E. & Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell*, **11**, 495–506.

Feuillet, C. & Keller, B. (2002) Comparative genomics in the grass family: molecular characterisation of grass genome structure and evolution. *Annals of Botany*, **89**, 3–10.

Foster-Hartnett, D., Mudge, J., Larsen, D., *et al.* (2002) Comparative genome analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G. *Genome*, **45**, 634–645.

Freeling, M. (2001) Grasses as a single genetic unit. *Plant Physiology*, **125**, 1191–1197.

Gale, M.D. & Devos, K.D. (1998) Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences, USA*, **95**, 1971–1974.

Gale, M.D., Moore, G. & Devos, K. (2001) Rice – the pivotal genome in cereal comparative genetics. In: *Rice Biotechnology: improving yield, stress tolerance and grain quality.* Novartis Foundation Symposium 236, pp. 46–58. John Wiley & Sons Ltd, New York.

Goff, S.A., Ricke, D., Lan, T.H., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science*, **296**, 92–100.

Glaszmann, J.C., Dufour, P., Grivet, L., *et al.* (1997) Comparative genome analysis between several tropical grasses. *Euphytica*, **96**, 13–21.

Grant, M.R., Godiard, L., Straube, E., *et al.* (1995) Structure of the *Arabidopsis RPM1* gene enabling dual-specificity disease resistance. *Science*, **269**, 843–846.

Grant, M.R., McDowell, J.M., Sharpe, A.G., *et al.* (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis. Proceedings of the National Academy of Sciences, USA*, **95**, 15843–15848.

Grant, D., Cregan, P. & Shoemaker, R.C. (2000) Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis. Proceedings of the National Academy of Sciences, USA*, **97**, 4168–4173.

Gualtieri, G., Kulikova, O., Limpens, E., *et al.* (2002) Microsynteny between pea and *Medicago truncatula* in the *SYM2* region. *Plant Molecular Biology*, **50**, 225–235.

Han, F., Kilian, A., Chen, J.P., *et al.* (1999) Sequence analysis of a rice BAC covering the syntenous barley *RPG1* region. *Genome*, **42**, 1071–1076.

Helentjaris, T., Weber, D.L. & Wright, S. (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics*, **118**, 353–363.

Heslop-Harrison, J.S. (2000) Comparative genome organisation in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell*, **12**, 617–635.

Hoenecke, M. & Chyi, Y.-S. (1991) Comparison of *Brassica napus* and *B. rapa* genomes based on restriction fragment length polymorphism mapping. In: Proceedings of the 8th International Rapeseed Congress GCIRC, Saskatoon (ed. D.I. McGregor), pp. 1102–1107.

Horimoto, K., Fukuchi, S. & Mori, K. (2001) Comprehensive comparison between locations of orthologous genes on archaeal and bacterial genomes. *Bioinformatics*, **17**, 791–802.

Hulbert, S.H., Richter, T.E., Axtell, J.D. & Bennetzen, J.L. (1990) Genetic mapping and characterisation of sorghum and related crops by means of maize DNA probes. *Proceedings of the National Academy of Sciences, USA*, **87**, 4251–4255.

Humphry, M.E., Konduri, V., Lambrides, C.J., *et al.* (2002) Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. *Theoretical and Applied Genetics*, **105**, 160–166.

Ito, M., Miyamoto, J., Mori, Y., *et al.* (2000) Genome and chromosome dimensions of *Lotus japonicus. Journal of Plant Research*, **113**, 435–442.

Ivanic, V., Malyshev, S., Korzun, V., Graner, A. & Børner, A (1999) Comparative mapping of a gibberellin acid-sensitive dwarfing gene (*DWF2*) on chromosome 4HS in barley. *Theoretical and Applied Genetics*, **98**, 728–731.

Jaffe, B., Caligari, P.D.S. & Snape, J.W. (2000) A skeletal linkage map of *Hordeum bulbosum* L. and comparative mapping with barley (*H. vulgare* L.). *Euphytica*, **115**, 115–120.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. & Dean, C. (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science*, **290**, 344–347.

Journet, E.P., Carreau, V., Gouzy, J., *et al.* (2001) The model legume *Medicago truncatula*: recent advances and perspectives in genomics. *OCL-Oleagineux Corps Gras Lipides*, **8**, 478–484.

Kaga, A., Ishii, T., Tsukimoto, K., Tokoro, E. & Kamijima, O. (2000) Comparative molecular mapping in *Ceratotropis* species using an interspecific cross between azuki bean (*Vigna angularis*) and rice bean (*V. umbellata*). *Theoretical and Applied Genetics*, **100**, 207–213.

Kawasaki, S. & Murakami, Y. (2000) Genome analysis of *Lotus japonicus. Journal of Plant Research*, **113**, 497–506.

Kennard, W., Phillips, R., Porter, R., Glombacher, A. & Phillips, R.L. (1999) A comparative map of wild rice (*Zizania palustris* L. 2n=2x=30). *Theoretical and Applied Genetics*, **99**, 793–799.

Kilian, A., Kudrna, D.A., Kleinhofs, A., *et al.* (1995) Rice-barley synteny and its application to saturation mapping of the barley *rpg1* region. *Nucleic Acids Research*, **23**, 2729–2733.

Kilian, A., Chen, J., Han, F., Steffenson, B. & Kleinhoffs, A. (1997) Towards map-based cloning of the barley stem rust resistance genes *rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. *Plant Molecular Biology*, **35**, 187–195.

Koch, M.A., Haubold, B. & Mitchell-Olds, T. (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Molecular Biology of Evolution*, **17**, 1483–1498.

Kole, C., Quijada, P., Michaels, S.D., Amasino, R.M. & Osborn, T.C. (2001) Evidence for homology of flowering time genes *VRF2* from *Brassica rapa* and FLC from *Arabidopsis thaliana*. *Theoretical and Applied Genetics*, **102**, 425–430.

Korzun, V., Røder, M., Worland, A.J. & Børner, A. (1997) Intrachromosomal mapping of genes for dwarfing (*RHT12*) and vernalization response (*VRN1*) in wheat by using RFLP and microsatellite markers. *Plant Breeder*, **116**, 227–232.

Kowalski, S.P., Lan, T.-H., Feldmann, K.A. & Paterson, A.H. (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organisation. *Genetics*, **138**, 499–510.

Ku, H.-M., Vision, T., Liu, J. & Tanksley, S.D. (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences, USA*, **97**, 9121–9126.

Kunkel, B.N. (1996) A useful weed to put to work – genetic analysis of disease resistance in *Arabidopsis thaliana*. *Trends in Genetics*, **12**, 63–69.

Kurata, N., Moore, G., Nagamura, Y., *et al*. (1994) Conservation of genome structure between rice and wheat. *Bio/Technology (NY)*, **12**, 276–278.

Lagercrantz, U. (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**, 1217–1228.

Lagercrantz, U. & Axelson, T. (2000) Rapid evolution of the family of *CONSTANS LIKE* genes in plants. *Molecular Biology in Evolution*, **17**, 1499–1507.

Lagercrantz, U. & Lydiate, D.J. (1996) Comparative genome mapping in *Brassica*. *Genetics*, **144**, 1903–1910.

Lagercrantz, U., Putterill, J., Coupland, G. & Lydiate, D. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale collinearity and congruence of genes controlling flowering time. *Plant Journal*, **9**, 13–20.

Lan, T.H., Delmonte, T.A., Reischmann, K.P., *et al*. (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Research*, **10**, 776–768.

Lee, J.M., Grant, D., Vallejos, C.E. & Shoemaker, R.C. (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. *Theoretical and Applied Genetics*, **103**, 765–773.

Leister, D., Kurth, J., Laurie, D.A., *et al*. (1998) Rapid reorganisation of resistance gene homologues in cereal genomes. *Proceedings of the National Academy of Sciences, USA*, **95**, 370–375.

Lydiate, D., Sharpe, A., Lagercrantz, U. & Parkin, I. (1993) Mapping the *Brassica* genome. *Outlook in Agriculture*, **2**, 85–92.

Lynch, M. & Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155.

Marek, L.F., Mudge, J., Darnielle, L., *et al*. (2001) Soybean genomic survey: BAC-end sequences near RFLP and SSR markers. *Genome*, **44**, 572–581.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., *et al*. (1993) Map-based cloning of a protein-kinase gene conferring disease resistance in tomato. *Science*, **262**, 1432–1436.

Mayer, K., Schuller, C., Wambutt, R., *et al*. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature*, **402**, 769.

Mayer, K., Murphy, G., Tarchinin, R., *et al*. (2001) Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Research*, **11**, 1167–1174.

Menancio-Hautea, D., Fatokun, C.A., Kumar, L., Danesh, D. & Young, N.D. (1993) Comparative genome analysis of mungbean (*Vigna radiata* L. Wilcek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. *Theoretical and Applied Genetics*, **86**, 797–810.

Michaels, S.D. & Amasino, R.M. (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, **11**, 949–956.

Michelmore, R. (2000) Genomic approaches to plant disease resistance. *Current Opinion in Plant Biology*, **3**, 125–131.

Moore, G., Devos, K.M. & Gale, M.D. (1995) Grasses, line up and form a circle. *Current Biology*, **5**, 737–739.

Mural, R.J., Adams, M.D., Myers, E.W., *et al.* (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science*, **296**, 1661–1671.

O'Brien, S.J. & Stanyon, R. (1999) Ancestral primate reviewed. *Nature*, **402**, 365–366.

O'Neill, C.M. & Bancroft, I. (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant Journal*, **23**, 233–243.

Oldroyd, G.E.D. & Staskawicz, B.J. (1998) Genetically engineered disease resistance in tomato. *Proceedings of the National Academy of Sciences, USA*, **95**, 10300–10305.

Osborn, T.C., Kole, C., Parkin, I.A.P., *et al.* (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus and Arabidopsis thaliana*. *Genetics*, **146**, 1123–1129.

Paterson, A.H., Lin, Y.R., Li, Z., *et al.* (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**, 1714–1718.

Paterson, A.H., Lan, T.-H., Reischmann, K.P., *et al.* (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics*, **14**, 380–382.

Paterson, A.H., Bowers, J.E., Burow, M.D., *et al.* (2000) Comparative genomics of plant chromosomes. *Plant Cell*, **12**, 1523–1539.

Peng, J.R., Richards, D.E., Hartley, N.M., *et al.* (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400**, 256–261.

Rossberg, M., Theres, K., Acarkan, A., *et al.* (2001) Comparative sequence analysis reveals extensive microcolinearity in the *lateral suppressor* regions of the tomato, *Arabidopsis* and *Capsella* genomes. *Plant Cell*, **13**, 979–988.

Sadowski, J. & Quiros, C.F. (1998) Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the RPS2 gene in the *Brassica nigra* genome. *Theoretical and Applied Genetics*, **96**, 468–474.

Sadowski, J., Gaubier, P., Delseny, M. & Quiros, C.F. (1996) Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Molecular and General Genetics*, **251**, 298–306.

Sasaki, A., Ashikari, M., Ueguchi-Tanak, M., *et al.* (2002) Green revolution: a mutant gibberellin synthesis gene in rice – New insight into the rice variant that helped to avert famine over thirty years ago. *Nature*, **400**, 256–261.

Schmidt, R. (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Molecular Biology*, **48**, 31–47.

Schmidt, R., Acarkan, A. & Boivin, K. (2001) Comparative structural genomics in the Brassicaceae family. *Plant Physiology and Biochemistry*, **39**, 253–262.

Shimamoto, K. & Kyozuka, J. (2002) Rice as a model for comparative genomics of plants. *Annual Review of Plant Biology*, **53**, 399–419.

Simpson, G.G. & Dean, C. (2002) Flowering – *Arabidopsis*, the rosetta stone of flowering time? *Science*, **296**, 285–289.

Slocum, M.K., Figdore, S.S., Kennard, W.C., Suzuki, J.Y. & Osborn, T.C. (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. *Theoretical and Applied Genetics*, **80**, 57–64.

Smith, S.F., Snell, P., Gruetzner, F., *et al.* (2002) Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and Human 20q. *Genome Research*, **12**, 776–784.

Snape, J.W., Sarma, R., Quarrie, S.A., Fish, L., Galiba, G. & Sutka, J. (2001) Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. *Euphytica*, **120**, 309–315.

Song, K.M., Suzuki, J.Y., Slocum, M.K., Williams, P.H. & Osborn, T.C. (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. *Theoretical and Applied Genetics*, **82**, 296–304.

Stougaard, J. (2001) Genetics and genomics of root symbiosis. *Current Opinion in Plant Biology*, **4**, 328–335.

Tabata, S., Kaneko, T., Nakamura, Y., *et al*. (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature*, **408**, 823–826.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. & Martin, G.B. (1999) Overexpression of *Pto* activates defence responses and confers broad resistance. *Plant Cell*, **11**, 15–30.

Tanksley, S.D., Bernatsky, R., Lapitan, N.L. & Prince, J.P. (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Sciences, USA*, **85**, 6419–6423.

Teutonico, R.A. & Osborn, T.C. (1994) Mapping of RFLP and quantitative trait loci in *Brassica rapa* and comparison to linkage maps of *B. napus*, *B. oleracea* and *Arabidopsis thaliana*. *Theoretical and Applied Genetics*, **89**, 7409–7414.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L. & Avramova, Z. (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proceedings of the National Academy of Sciences, USA*, **96**, 7409–7414.

Torres, A.M., Weeden, N.F. & Martin, A. (1993) Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. *Theoretical and Applied Genetics*, **85**, 937–945.

Truco, M.J. & Quiros, C.F. (1994) Structure and organization of the B genome based on a linkage map in *Brassica nigra*. *Theoretical and Applied Genetics*, **89**, 590–598.

Truco, M.J., Hu, J., Sadowski, J. & Quiros, C.F. (1996) Inter- and intra-genomic homology of the *Brassica* genomes: implications for their origin and evolution. *Theoretical and Applied Genetics*, **93**, 1225–1233.

van der Beizen, E.A. & Jones, J.D.G. (1998) Plant disease resistance proteins and the gene-for-gene concept. *Trends in Biochemical Science*, **23**, 454–456.

van Deynze, A.E., Nelson, J.C., Yglesias, E.S., *et al.* (1995a) Comparative mapping in grasses. Wheat relationships. *Molecular and General Genetics*, **248**, 744–754.

van Deynze, A.E., Nelson, J.C., O'Donoughue, L.S., *et al.* (1995b) Comparative mapping in grasses. Oat relationships. *Molecular and General Genetics*, **249**, 349–356.

Vision, T.J., Brown, D.G. & Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.

Weeden, N.L., Muehlbauer, F.J. & Ladizinsky, G. (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of Heredity*, **83**, 123–129.

Wroblewski, T., Coulibaly, S., Sadowski, J. & Quiros, C.F. (2000) Variation and phylogenetic utility of the *Arabidopsis thaliana RPS2* homolog in various species of the tribe Brassiceae. *Molecular and Phylogenetic Evolution*, **16**, 440–448.

Yang, Y.W., Lai, K.N., Tai, P.Y. & Li, W.H. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *Journal of Molecular Evolution*, **48**, 597–604.

Yu, J., Hu, S., Wang, J., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.

4 Plant genetic engineering

Ian Puddephat

4.1 Background

Genetic transformation of plant species is a naturally occurring phenomenon (Hooykaas & Schilperoort 1992). Our ability specifically to transfer individual genes, singly or in defined sets, stems from transformation experiments first described in tobacco (Barton *et al.* 1983; Herrera-Estrella *et al.* 1983; Deblock *et al.* 1984, Horsch *et al.* 1984). The most commonly used transformation methods to transfer DNA to plant cells either exploit the natural capacity of the bacterium *Agrobacterium tumefaciens*, or use microprojectiles coated in DNA, that are 'fired' into plant cells. Gene transfer methods are usually combined with suitable tissue culture techniques for regenerating whole fertile plants.

This chapter provides a brief review of the principal plant transformation technologies available for gene transfer to plant cells and the variables that affect efficiency of transformation. The purpose is to identify the key problems and practical limitations of these approaches and to review current research efforts aimed at increasing the efficiency of transformation. Emphasis is given to a few key areas. These are, first, new or novel methods of transformation aimed at increasing the efficiency with which genes can be tested, expressed or contained following release of transgenic material. Second, methods to extend the utility of transgenic material by seeking to remove DNA sequences used only to mark or select transformation events. Third, consideration is given to current restrictions in the adoption of transformation technologies to a broader range of genotypes. There is also a discussion of research that is seeking to identify plant genes that regulate DNA transfer potentially allowing increased efficiency of gene transfer and integration, and possibly a better means of targeting genes to specific regions of the genome.

A series of authors has reviewed progress in plant transformation, and their publications serve to chart the development of the key technologies (Potrykus 1991; Zambryski 1992; Kikkert 1993; Sanford *et al.* 1993; Songstad *et al.* 1995; Zupan & Zambryski 1995; Christou 1996; Birch 1997; Gelvin 1998; Hansen & Wright 1999). Initially, emphasis focused on gene transfer technologies and strategies for creating transgenic crops (see Walden & Wingender 1995). A wealth of literature has been accumulated during the past two decades and significant progress in the development of gene transfer technologies has been made, such that the first major plantings of genetically engineered (or transgenic) crops took place in 1996 (James 1997). In the same year Christou (1996) reviewed the state-of-the-art of transformation

technologies, concluding that is was remarkable 'that in just over a decade the tools of recombinant DNA technology and cell biology' were already 'at the disposal of plant breeders'.

Published protocols for the transformation of well over 120 plant species are available (Birch 1997; Hansen & Wright 1999), and these include most major economic crops, vegetables, ornamental, medicinal, fruit, tree and pasture plants. The pace at which the technology has advanced makes listing transformation techniques developed for each species futile, and information for a particular species is best gained through searches of bibliographic or patent databases. Transformation technologies have been successfully combined within plant breeding programmes for more efficient introduction of genes or the generation of plants with phenotypes unachievable by conventional plant improvement. To date, regulatory approvals have been granted for the transgenic modification of cultivars of 16 species and the global area under cultivation with genetically modified (GM) crops exceeded 55 million hectares in 2001 (James 2001).

Although transformation technologies have developed quickly and are being deployed commercially, GM products must also be received favourably by society and command a place in the market. To this end, regulations are in place concerning the gene(s), the regulatory sequences used to drive them, and any other sequences used to mark or select transformation events. The use of alternative markers to genes conferring resistance to antibiotics and herbicides has become increasingly important to satisfy regulatory concerns (Royal Society 1998). Several alternative marker genes have become available, as have strategies to remove them from the final transgenic line. Marker-free transformation may well extend the utility of transgenic lines in plant breeding, allowing the adoption of breeding strategies previously unattainable (e.g. Pink & Puddephat 1999). Several GM products have commanded places in the market already and have proved highly successful with both producers and consumers. Reactions to future advances are less easy to predict in the current climate. Technological advances are likely that increase our ability to produce genetically engineered plants with greater precision, and these will address some of the current concerns. The benefits of GM crop plants and products to both the environment and consumer will need to be more transparent – a difficult task given the complexities of the agricultural environment and market place.

Currently approved GM plants have been developed to express one or more of the following basic phenotypic characteristics: delayed ripening; fertility restoration; herbicide tolerance; insect resistance; male sterility; modified colour; modified oil; and virus resistance. Productivity gains have already been achieved through genetic modifications for herbicide tolerance and resistance to viral pathogens and insect pests. Current genomic research offers enormous potential for further improving production-related traits. However, the ability to manipulate specific genes through transgenesis has itself provided a powerful tool with which to study the regulation of physiological and developmental processes (Newbigin *et al.* 1995), and this has led to the discovery of large numbers of genes that represent further targets for the genetic engineering of plant traits. Such traits include apical dominance (Medford

et al. 1989), the regulation of floral development (Newbigin *et al.* 1995; Williams 1995) and pigmentation (Mol *et al.* 1995), plant–microbe interactions (Salmeron & Vernooij 1998), senescence (Chandlee 2001) and fruit ripening (Theologis *et al.* 1992; Lelievre *et al.* 1997), photosynthesis (Paul & Foyer 2001), stress (Van Breusegem *et al.* 1998) and the regulation of other metabolic pathways (Herbers & Sonnewald 1996, 1998). This supply of target genes has been enormously enhanced by the current explosion in genomics programmes, facilitated by advances in plant molecular biology and the sequencing of the *Arabidopsis* and rice genomes. New opportunities are arising to test the function of a diverse range of genes or tailor their expression/activity for specific improvement programmes. As well as the possibility of manipulating the type of traits that have long been the targets of plant breeders, transformation technologies also offer opportunities for the introduction of novel genes to create 'plant factories' for the production of pharmaceuticals, nutraceuticals and other high-value proteins or chemicals (Alonso & Maroto 2000; Fischer & Emans 2000; Fischer *et al.* 2001).

4.1.1 Key components of plant genetic engineering

The successful production of transgenic plants has four essential requirements. These are suitable vectors, selectable markers (both covered in detail later) and efficient techniques for both the transformation and regeneration of whole plants. For most transformation systems, a tissue culture stage is a prerequisite for the recovery of whole plants. Advances in plant tissue culture have certainly resulted in significant improvements in transformation for several species (Hansen & Wright 1999), but there are no universally applicable methods. Rather, considerable genotypic variation is frequently encountered in the development of transformation and regeneration protocols and is frequently a major restriction in the application of the technology. The efficiency of transformation is often limited by an inability to transform cells that are competent for regeneration, as the two processes can occur in independent cell types (Colby *et al.* 1991; Jordan & McHughen 1988). Whilst there has been some progress in developing understanding of competency for transformation at the cellular level (Sangwan *et al.* 1992; Villemont *et al.* 1997), we are only just beginning to identify the plant genes that participate in the process.

Several authors have sought to define plant transformation by providing criteria upon which the success of gene transfer technologies can be judged (Potrykus 1991; Jenes *et al.* 1993; Birch 1997). More recently, fluorescence in-situ hybridization (FISH) has been used to localize transgenes on chromosomes of several plant species (Fransz *et al.* 1996; Hoopen *et al.* 1996; Pedersen *et al.* 1997). The method allows detection of copy number and the assessment of integration sites, providing insights into the effect of transgene position on rates and stability of expression (Kohli *et al.* 1998; Hoopen *et al.* 1999). Recent advances have provided methods for the detection of regions of DNA less than 1 kb in length by FISH (Salvo-Garrido *et al.* 2001; Khrustaleva & Kik 2001).

Phenotypic data showing expression of the introduced gene in lines confirmed to carry the gene(s) by Southern analysis, provide further evidence for stable integration of transgenes. However, ambiguous results can be obtained if appropriate tests and controls are not employed. Controls should be generated using the same methods used to generate the transgenic lines. Assays should be based upon the detection of genes that are not expressed within bacteria used as vectors, for which purpose intron-containing genes such as *GUS* and *NPT*II or anthocyanin reporter genes are suitable. Further evidence for transformation is obtained when the transmission of an introduced gene(s) and the resulting phenotype can be confirmed in sexual offspring (Potrykus 1991). However, as Birch (1997) pointed out, this is not always an option where the target plant for transformation is either long-lived (e.g. forest tree species) or propagated vegetatively (e.g. strawberry, potato, banana).

4.1.2 Methods for plant transformation

The technology of plant transformation has progressed rapidly, and there are excellent reviews documenting the development of *Agrobacterium*-mediated transformation (Gelvin 1990, 2000; Kado 1991; van Wordragen & Dons 1992; Zupan *et al.* 2000), direct gene transfer into protoplasts (Songstad *et al.* 1995) and particle bombardment (Sanford *et al.* 1993; Southgate *et al.* 1995). The potential practical application of these technologies have been considered by several authors (Vandenelzen *et al.* 1993; Dale & Irwin 1995; Lydiate *et al.* 1995; Kramer & Muthukrishnan 1997; Shah 1997; Evans & Greenland 1998; Salmeron & Vernooij 1998; Pink & Puddephat 1999). Similarly, comprehensive descriptions and discussion of transformation methodologies can be found in various texts (Walden 1988; Grierson 1991; Kung & Wu 1993) and method manuals (Draper *et al.* 1988; Gelvin & Schilperoort 1988; Croy 1993; Potrykus & Spangenberg 1995).

In order successfully to transform a plant, certain criteria must be satisfied. These can vary widely depending on the application, but a useful set identified by Hansen and Wright (1999) include:

- A target tissue that is competent for regeneration and/or propagation.
- An efficient means of delivering DNA.
- Agents to either select or identify transgenic tissues.
- The ability to recover fertile transgenic plants at a reasonable frequency.
- A simple, efficient, reproducible, genotype-independent and cost-effective process.
- A short time-frame in culture to avoid somaclonal variation.

In the course of product development, the structure and copy number of transgenes along with their stability must also be established, and this can have a bearing upon the method chosen for transforming a crop species. In most situations the ideal transgenic line would possess a single simple insertion event with no extraneous DNA from the plasmid vector employed to effect DNA transfer. In this context, the various

methods of transformation have both advantages and disadvantages. Techniques that appear to fulfil the criteria set out above include protoplast transformation, microprojectile bombardment (also referred to as biolistics) and *Agrobacterium*-mediated transformation. Various additional methods for gene transfer to plants have been developed; some have specific applications but may not meet all the criteria for development of a successful transformation system. No single approach has yet to prove effective in all species and it is likely that a range of gene transfer technologies will continue to be developed and refined.

The methods for introducing DNA into plant cells are conveniently divided into two types: direct, and indirect. Methods based on the use of *Agrobacterium* or viruses fall into the latter category, where a vector is used to transfer DNA into cells of the target plant. Direct delivery of DNA is achieved using microprojectile bombardment, electroporation, microinjection or use of silicon whiskers (Hansen & Wright 1999). There has, however, been some blurring of these distinctions, most notably the combining of elements of *Agrobacterium*-mediated transformation with microprojectile bombardment (Bidney *et al.* 1992; Hansen & Chilton 1996) and microinjection (Escudero *et al.* 1995, 1996). By far the most successful approaches have been those based upon *Agrobacterium*-mediated transfer and microprojectile bombardment optimized to deliver genes into the nuclear genomes of a range of plant species. Techniques for the genetic engineering of the plastid genome have been developed (Svab *et al.* 1990) following optimization of the selection process (Svab & Maliga 1993), and these methods are becoming increasingly important (Maliga 2001; Ruf *et al.* 2001).

4.2 *Agrobacterium*-mediated transformation

Agrobacterium tumefaciens elicits tumour growths on a range of plant species by genetically engineering cells with a specific segment of its own DNA. This ability led to the development of *Agrobacterium* as a vector for genetic engineering of flowering plants (Herrera-Estrella *et al.* 1983). The biological process by which *Agrobacterium* elicits transformation events has been the subject of intensive research which continues to try and improve our understanding of the transfer mechanism. The process involves a complex series of chemical signals between the bacterium and the host that result in the movement of a segment of T(transfer)-DNA from the bacterium into the host plant cells and its subsequent integration into the nuclear genome. Evidence indicates that the mechanism of T-DNA transfer is functionally and evolutionarily related to bacterial conjugation (Christie & Vogel 2000; Zupan *et al.* 2000). It involves expression of a set of *vir* (virulence) genes (Stachel & Nester 1986) found largely on a large extra-chromosomal plasmid (~200 kb) but also in the bacterial chromosome (Zupan & Zambryski 1997; see also the review of Kado 1991). The expression of *vir* genes is optimal at acidic pH and temperatures of 19 to 22°C (Fullner *et al.* 1996; Fullner & Nester 1996) in the presence of phenolic compounds (e.g. acetosyringone) that are normally involved in plant phytoalexin and lignin

biosynthesis, and are released following damage to plant cells. Induction of *vir* gene expression by plant phenolics (Winans 1991) results in attachment of the bacterium to the host cell and the formation and export of the T-DNA (for review, see Zupan & Zambryski 1997). A great deal is known about the mechanisms by which *Agrobacterium* delivers DNA to plant genomes (Hooykaas & Beijersbergen 1994; Kado 1994; Zupan & Zambryski 1995; Sheng & Citovsky 1996; Kumpatla *et al.* 1998; Zupan *et al.* 2000), but much less is known of the role of host genes and proteins in the process (Gelvin 2000).

Agrobacterium-mediated transformation is an attractive system. At a technical level, the process is simpler than direct gene transfer methods and has minimal equipment costs. Experience has shown that plant cells transformed using direct gene transfer technologies bring about insertion of unwanted DNA, multiple and fragmented copies at high copy numbers that can cause expression instability. In general terms, *Agrobacterium*-mediated transformation produces plants with single (or simple) insertions of low copy number and few re-arrangements (Chilton 1993; Gelvin 1998). However, *Agrobacterium*-mediated gene transfer is not without problems. Superfluous DNA transfer (Smith *et al.* 2001) and persistence of the bacterium with primary transformants (Matzk *et al.* 1996) create problems for the release of transgenic plants in some species. Nevertheless, the advantages of *Agrobacterium*-mediated transformation have frequently been the driving force behind studies to develop protocols for new or recalcitrant species such as the monocotyledonous crop plants.

Typically *Agrobacterium*-mediated transformation is combined with a plant tissue culture procedure and is performed by inoculating plant material with *Agrobacterium* followed by a period of co-cultivation. Protocols have been developed for inoculation of cells (including protoplasts), roots, hypocotyls, cotyledons, stem sections, leaf pieces and germinating seeds. The process of *Agrobacterium*-mediated transformation requires optimization of two biological systems. It has been common practice to first establish gene transfer to plant cells using *Agrobacterium* and then develop a means of regeneration (van Wordragen & Dons 1992). As already stated, in many systems cells that are competent for transformation may not necessarily participate in regeneration (Colby *et al.* 1991). In these circumstances, Birch (1997) recommends using histological studies to determine the pattern of events and cellular origins of regenerating structures. A number of parameters can be adjusted in the development of a protocol and a useful account of the approach has been provided by van Wordragen and Dons (1992).

Initially, protocol development starts with selection of virulent bacterial strains and a responsive plant cultivar or variety; significant interactions between strain and plant genotype occur (Mauro *et al.* 1995). Choice of bacterial strain can significantly affect chances of success; for example, little difference in the virulence of *A. tumefaciens* strains on *Brassica juncea* was found, whereas on *B. napus* only nopaline strains were efficient (Charest *et al.* 1989). Similar differences have been reported in soybean (Byrne *et al.* 1987) and in general terms nopaline strains of *A. tumefaciens* are more virulent (van Wordragen and Dons, 1992).

An understanding of the molecular mechanisms of T-DNA transfer has allowed development of several approaches to enhance the virulence of the bacterial strain used. Treatments to enhance virulence are either applied directly to the bacteria or result following manipulation of the target plant tissue. Pre-induction or enhanced induction of the bacterial *vir* genes can be achieved through the use of acetosyringone and related compounds. Such approaches are only of value where the target tissue produces insufficient phenolic compounds to induce *vir* gene expression, which may also occur where short co-cultivation periods are used. Increasing the level of tissue wounding, which increases the release of phenolic compounds, can stimulate bacterial virulence and provide greater access to transformable tissues although pre-induced bacteria can affect transformation in the absence of cell injury (Escudero & Hohn 1997). Pre-wounding tissues with either glass beads or microprojectiles has proven effective (Grayburn & Vick 1995; May *et al.* 1995), as has the use of infiltration or ultrasound treatment, through sonication (Santarem *et al.* 1998). Increasing transformation efficiency has also been demonstrated using feeder or nurse layers, comprising suspensions of cells from readily transformable species such as petunia, potato or tobacco. Target tissue for transformation is usually cultured on the feeder layer separated by a filter paper. The beneficial effects of this approach may be due to secretion of phenolic compounds stimulating *vir* gene expression, or other compounds released that influence morphogenesis, or a reduction of the negative effects of tissue culture or bacterial inoculation.

Bacterial virulence is also affected by environmental conditions. *vir* gene induction has been reported to be optimal at acidic pH (Altmorbe *et al.* 1989), and in several species shifts in pH can have pronounced effects on the efficiency of transformation (Godwin *et al.* 1991). Optimal temperatures have been variously reported to be below 28°C for *vir* gene induction (Altmorbe *et al.* 1989), 19°C for T-DNA transfer (Fullner & Nester 1996; Fullner *et al.* 1996), 19 to 22°C for T-DNA transfer and transformation in several species (Dillen *et al.* 1997), and 25°C for obtaining the highest number of stable transformants in tobacco (Salas *et al.* 2001). Such observations reinforce the importance of considering *Agrobacterium*-mediated transformation as a two-component system. Whilst optimal T-DNA transfer occurs at a lower temperature, the evidence of Salas *et al.* (2001) indicates that co-cultivation at a higher temperature can be beneficial for plant cell susceptibility to infection and stable T-DNA insertion into the plant genome.

As has been indicated, manipulation of bacterial virulence is not the only route to achieving increases in the efficiency of *Agrobacterium*-mediated transformation. In several crops, plant tissues appear to be sensitive to inoculation with *Agrobacterium*. Necrotic barriers are formed in grape tissues that serve to restrict efficiency of transformation. Treatment with antioxidants (e.g. polyvinylpyrrolidone and dithiothreitol) inhibited tissue necrosis without affecting *Agrobacterium* virulence, thereby enabling the recovery of stable transgenic grape plants resistant to hygromycin (Perl *et al.* 1996). Similar necrotic reactions have been noted in *Brassica* species (Ohlsson & Eriksson 1988); these are frequently related to the physiological condition of the starting material and can also be overcome by frequent rinsing of inoculate

tissues to reduce bacterial populations (Trail *et al.* 1989). There is some evidence to indicate that Agrobacteria do not induce hypersensitive reactions on plant species (Robinette & Matthsse 1990); nevertheless there are pronounced differences in the transformability of even closely related genotypes and there exists the possibility that these genotypic differences may be related to various plant resistance mechanisms. Agrobacteria interact with a range of plant-encoded proteins during all steps of transformation (see Section 4.8.2), including at the initial step of bacterial attachment to the plant cell. Differences in the susceptibility of plant genotypes to *Agrobacterium*-mediated transformation have been noted to occur at attachment in certain explant types but not in other tissues of the same genotype (Nam *et al.* 1997). Consequently, manipulations aimed at overcoming deficiencies within the bacteria may well prove to be limited and in order to resolve fully the poor frequencies of transformation a greater understanding of the plant genes/proteins involved in the process will be required.

4.2.1 Agrobacterium rhizogenes

Natural mechanisms of gene transfer to plant cells also operate in the induction of hairy root disease by *Agrobacterium rhizogenes.* Virulent strains of *A. rhizogenes* carry a Ri (root inducing) plasmid that shares key characteristics with Ti plasmids: a T-DNA and a virulence region. The virulence genes of Ri and Ti plasmids are very similar and can be interchanged (Hooykaas *et al.* 1984; Gelvin 1990). Although the mechanisms of gene transfer are the same, the physiological basis of hairy root disease and crown gall tumours are very different (Gelvin 1990). *A. rhizogenes* induces hairy roots on a range of dicotyledenous species (see Tepfer 1990), and transgenic plants have been regenerated from hairy roots of some of these species (see Christey 1997, 2001). As with *A. tumefaciens*, *A. rhizogenes* is not a natural host of monocotyledonous plants, but gene transfer using *A. rhizogenes* has been reported in onion (Dommisse *et al.* 1990). The use of *A. rhizogenes* for production of transgenic plants has recently been reviewed (Christey 2001).

Phenotypic alterations are frequently found in transgenic plants regenerated from hairy roots. These are principally caused by expression of the *rol* genes carried on the Transfer Left (TL)-DNA transferred from Ri plasmids, although other genes may also be involved. One of the obvious alterations is the increase in rooting – a characteristic that has been used to advantage in several species to overcome restrictions in vegetative propagation (see Christey 1997). Morphological alterations in *A. rhizogenes*-transformed plants are common, including reduced internode elongation, wrinkled leaves and increased branching. Changes of both perennial and biennial plants to annual habits have also been observed following *A. rhizogenes* transformation. Flowering is inhibited in some species, and altered flower morphology and male sterility have all been reported. Some of these alterations in morphology are specific to particular *rol* genes. The expression of *rol*C has been shown to lead to male sterility (Schmulling & Schell 1993). Expression of *rol*A and *rol*C is known to modify shoot morphology, leading to dwarfing and severely wrinkled leaves (Schmulling *et*

al. 1988). Increased root formation is observed following expression of *rol*B, which also causes reduced apical dominance and internode length.

Gene transfer using *A. rhizogenes* is attractive because a wide range of plant species can be regenerated from hairy roots either spontaneously or indirectly via a callus phase following transfer to a medium containing plant growth regulators. The altered phenotype obtained following *A. rhizogenes*-mediated transformation is frequently regarded as a disadvantage and has restricted its application in plant genetic engineering. Nevertheless, in flax, black locust and *Stylosanthes humilis*, *A. rhizogenes*-mediated transformation has proven to be an efficient alternative *to A. tumefaciens* (Zhan *et al.* 1988; Manners & Way 1989; Han *et al.* 1993). A feature of *A. rhizogenes*-mediated transformation is the speed with which transformed cell lines can be generated. In the present author's laboratory, transgenic hairy roots can be isolated from seedling explants of *Brassica* as proliferating cell lines within 3 to 4 weeks. Consequently, for simple assays of gene function – and particularly for genes encoding enzymes in metabolic pathways – the system can be adapted as a means of high-throughput gene testing.

In addition to the speed with which transformation events can be obtained following *A. rhizogenes*-mediated transformation, two further advantages are apparent. First, the hairy root morphology can be used as a means of selecting transgenic cell lines. Although this is not efficient in all species (Berthomieu & Jouanin 1992), it has recently been successfully combined with reporter genes (*gfp* and *gus*) to provide a versatile means of rapidly identifying transformation events (Cogan *et al.* 2001; Puddephat *et al.* 2001). Consequently, it is possible to obtain transgenic cells without the need for selection agents that may inhibit shoot regeneration, and thus selectable marker genes can be omitted. Second, in contrast to *A. tumefaciens*, there are no reports of the regeneration of chimeric plants following *A. rhizogenes*-mediated transformation. Following *A. tumefaciens*-mediated transformation, adventitious shoots are frequently regenerated in culture. Adventitious shoot regeneration requires the participation of several cells. If selection is inefficient, then mixtures of non-transformed and transformed cells may participate during regeneration resulting in the production of chimeric transgenic plants. Generally, these problems are addressed by adjusting selection strategies. However, such difficulties do not arise when *A. rhizogenes* strains are used as hairy roots appear to be derived from transformation events to single cells (Tempe & Casse-Delbart 1989). Proliferating hairy root cultures, established by root tip culture, have been shown to be composed of cells carrying identical transformation events, as have the transgenic plants regenerated from them. Furthermore, hairy root cultures exhibit a high degree of genetic and biochemical stability during long-term culture (Hamill & Lidgett 1997) and are consequently an efficient means of preserving transgenic cell lines for a range of purposes.

The strategies for gene transfer with *A. rhizogenes* differ from those with *A. tumefaciens*. It is not possible to regenerate plants from tumours induced by wild-type strains of *A. tumefaciens*, with one notable exception (Berthomieu *et al.* 1994). Consequently, disarmed strains of *A. tumefaciens* have been developed for plant trans-

formation, in which the T-DNA regions carrying the genes that cause tumorigenesis have been removed. For *A. rhizogenes*, the ability to obtain regeneration of plants from hairy roots negates the need for disarmed strains; indeed, preserving the ability to induce hairy roots allows selection of transformed material. Some of the vectors used for *A. rhizogenes*-based transformation are described in the next section.

Since the *vir* gene products from Ri plasmids of *A. rhizogenes* will also transfer T-DNA regions from binary plasmids *in trans* (see next section), it is possible to co-transfer T-DNAs from Ri plasmids and binary vectors resident in the same bacterial strain. High rates of co-transfer of Ri and binary plasmid T-DNAs have been reported in hairy roots of *N. tabacum*, *Brassica napus* and *Medicago arborea* (Boulter *et al.* 1990; Hatamoto *et al.* 1990; Damiani & Arcioni 1991). Where Ri and binary vector T-DNAs insert into the genome at independent genetic loci they will segregate in subsequent generations. This allows recovery of transgenic plants that are phenotypically unaltered (containing no Ri-plasmid T-DNA) as has been achieved in tobacco, rapeseed and cauliflower (Boulter *et al.* 1990; Hatamoto *et al.* 1990; Puddephat *et al.* 2001). This approach has great potential for the generation of marker-free transgenic lines (see later). As was demonstrated in cauliflower (Puddephat *et al.* 2001), co-transformed hairy roots could be selected without using chemical agents, allowing for the possibility of engineering binary vector T-DNAs with only genes of interest so that plants containing only these sequences can be recovered in the second generation. The approach has considerable merit, although its application is currently restricted to species in which the *rol* phenotype does not severely impair fertility. Much is known about the individual roles of the *rol* genes in root induction and the altered phenotype of Ri transformed plants. It is likely that fertility problems would be reduced if *A. rhizogenes* strains containing only *rol*B, known to induce hairy root formation in isolation, were engineered or strategies for eliminating other *rol* genes employed. Some progress has been made on the latter aspect but the potential of *A. rhizogenes*-mediated transformation has yet to be realized.

4.2.2 Monocot transformation

Initial attempts to transform monocotyledonous plants with *Agrobacterium* were largely unsuccessful. Monocotyledonous plants are not a natural host for *Agrobacterium* (Porter 1991) because their cells produce different phenolic compounds in response to wounding and respond differently to the hormonal cues that induce tumour growth in dicots (see Kado 1991). In some monocots, exudates from specific tissues have been found that are potent inhibitors of virulence gene induction (Zhang *et al.* 2000). However, the lack of tumour induction does not necessarily imply a lack of virulence on the part of the bacterium and it has frequently been shown that T-DNA transfer is not a rate-limiting step to transformation. Early results with maize cells infected with *Agrobacterium* showed that the initial kinetics of expression of the *gus* reporter gene are similar to those in tobacco. However, the presence of *gus* mRNA in maize is highly transient, suggesting that the block to maize transforma-

tion involves T-DNA integration rather than T-DNA entry into the cell or nuclear targeting (Narasimhulu *et al.* 1996).

Successful *Agrobacterium*-mediated transformation of monocotyledonous crop plants has largely resulted from the exploitation of our knowledge of the *Agrobacterium* genes controlling T-DNA transfer. In maize, several studies have shown that T-DNA transfer is much more effective from nopaline than octopine strains of *A. tumefaciens.* It seems that in this situation the VIRA protein of the nopaline strain is more effective for maize transformation than its homologue in octopine strains (reviewed in Chilton 1993). 'Super-binary plasmids' and 'super-virulent' bacterial strains, that either carry additional copies of *vir* genes or have enhanced *vir* gene induction, have been important for the transformation of several monocots, including wheat, maize and rice (Ishida *et al.* 1996; Guo *et al.* 1998; Khanna & Raina 1999; Ke *et al.* 2001). A recently developed set of plasmids allows integration of additional copies of *vir* genes, or other DNA sequences, into the chromosomal background of bacterial strains to enhance the efficiency of gene transfer (Lee *et al.* 2001). Pre-induction and co-cultivation of *Agrobacterium* in acetosyringone, to induce the virulence genes, has also been critical to transformation success, for example in rice (Khanna & Raina 1999) and wheat (Guo *et al.* 1998).

Utilization of *Agrobacterium* for transformation of monocots has also required advances in plant tissue culture technologies. Selection of appropriate starting material for *Agrobacterium*-mediated transformation of monocotyledonous plants is an important determinant of success (Nadolska-Orczyk & Orczyk 2000). Transformants of a maize inbred line A188 were first efficiently produced using immature embryos co-cultivated with *A. tumefaciens* strains harbouring 'super-binary' vectors (Ishida *et al.* 1996). Successful *Agrobacterium*-mediated transformation of rice initially required use of tissues consisting of actively dividing, embryonic cells, such as immature embryos and calli induced from scutella (Hiei *et al.* 1997), although more recently rice seedling root and shoot tips have been successfully transformed with super-virulent bacterial strains (Khanna & Raina 1999). Immature embryo cells have been transformed in barley and wheat (Cheng *et al.* 1997; Guo *et al.* 1998; Trifonova *et al.* 2001). Treatments aimed at limiting necrotic reactions in explants introduced into culture and co-cultivated with *Agrobacterium* have also proven effective in monocot species (Enriquez-Obregon *et al.* 1999).

A major attraction of developing *Agrobacterium*-mediated transformation for monocots is the 'clean and tidy' insertion pattern of T-DNA (Chilton 1993). Ishida *et al.* (1996) reported stable integration, expression, and inheritance of the transgenes in transgenic lines of maize, with between one and three copies. Transgenes were also found to be integrated with little rearrangement, and the boundaries of T-DNA were reported to be similar to those in transgenic dicotyledons. Similar results have been reported in wheat and rice (Cheng *et al.* 1997; Hiei *et al.* 1997). Efficient protocols for the regeneration of high-quality and fertile transgenic plants from transformed cells have been developed in monocots and the genotypic restrictions apparent in *Agrobacterium*-mediated transformation of dicots are also apparent here. For example, whilst transformation of the maize line A188 has been success-

ful, F_1 hybrids between A188 and five other inbreds are reported to transform at low frequencies (Ishida *et al.* 1996).

4.2.3 Vectors for Agrobacterium*-mediated transformation*

A considerable number of plant transformation vectors for *Agrobacterium*-mediated transformation have been developed. These are essentially either intermediate (or integrative) or binary (or autonomous) vectors and they share two principal features. First, transfer of the vector T-DNA to the plant is mediated by the ~35 bacterial virulence genes. Second, each vector contains a T-DNA region, which is defined by the right and left border sequences present in wild-type T-DNAs. The vectors have also been designed for ease of DNA manipulation in *Escherichia coli* prior to transfer to *Agrobacterium*. Essential features include the presence of unique multiple cloning sites (MCSs) for introducing target genes within the T-DNA region and a bacterially expressed marker gene that permits selection and maintenance of the vector in *E. coli* and *Agrobacterium*. A variety of bacterially expressed marker genes have been used, the most common being genes encoding resistance to kanamycin, ampicillin, tetracyclin or spectinomycin. Selection of the selectable marker for plasmid maintenance in bacteria requires care. It must not duplicate resistance functions encoded by genes residing within the bacterial chromosome or on other plasmids within the strain to be used, or provide resistance to antibiotics used to limit growth of the bacterium following inoculation and co-cultivation with plant cells. Antibiotic resistance genes frequently make up part of the T-DNA where they are used to select transformed plant cells. Since T-DNA transfer occurs from the right to left border, it is common practice to place the selectable marker near the left border to ensure transfer of the gene of interest. However, this was not the arrangement in many of the early vectors (e.g. pBIN19).

4.2.3.1 Integrative vectors

In the first type of vector, T-DNA is recombined via homologous sequences between the intermediate vector and the resident Ti plasmid of the bacterial strain (see, for example Draper *et al.* 1988; White 1993). Intermediate vectors are designed such that they cannot replicate in *Agrobacterium* without first integrating into the resident Ti plasmid. To allow this, intermediate vectors carry a region of homology to Ti plasmids which is usually a short segment of Ti plasmid DNA from the target site for integration and tends to be ~1 to 2 kb in size. Some vectors contain two homologous sequences that, following a double recombination event, ensure the sequences of the target site are not duplicated. The vector may be engineered with both a left border (LB) and a right border (RB), neither, or just a RB depending on whether limiting transfer of the backbone of the intermediate vector is important or not. Many intermediate vectors, used in both *A. rhizogenes* and *A. tumefaciens*, have been based around the replication functions of pBR322 (Rogers *et al.* 1986; Robaglia *et al.* 1987) that has a ColE1-type origin and is restricted to enteric bacteria such as *E. coli*. Intermediate vectors also carry functions, such as the *bom* site of pBR322,

which allow their mobilization from *E. coli* to *Agrobacterium*. To effect transfer of intermediate vectors to *Agrobacterium*, specific transfer functions are required. These are usually supplied *in trans* by a second plasmid usually resident in another bacterial strain. Consequently, mobilization of intermediate vectors usually requires three bacterial strains: the recipient *Agrobacterium*, an *E. coli* strain harbouring the intermediate and a further strain carrying the helper plasmid. Such transfers are referred to as triparental matings.

Several co-integrate vectors have been developed to introduce genes of interest into the T-DNA of Ri plasmids (Robaglia *et al.* 1987; Tempe & Casse-Delbart 1989). For example, Berthomieu and Jouanin (1992) used this approach to transfer hygromycin and kanamycin resistance genes, co-integrated into the T-DNA of the virulence plasmid pRiA4, to *Brassica oleracea*. The method has also been used to express pathogen-derived sequences in transgenic plants designed to provide resistance to viral infection (Zaccomer *et al.* 1993). A disadvantage of this approach is that the inserted genes of interest become physically linked to the *rol* genes and are thus expressed in plants with altered phenotypes.

4.2.3.2 Binary vectors

Binary or autonomous vectors differ from integrative vectors in that they can be stably maintained in *Agrobacterium*. They carry a copy of T-DNA that is mobilized into the plant genome by the *vir* genes present on a resident wild-type or disarmed Ti or Ri plasmid. Binary vectors are maintained in *Agrobacterium* through the use of broad host range replication origins carried on the plasmid. Many binary vectors utilise the replication origin of RK2, although this permits only maintenance of low copy numbers in *E. coli* and can be a limiting factor in DNA manipulation. The ColE1 origin of replication from pBR322 permits higher copy numbers in *E. coli* while other vectors have used the pSa or pRi origins. In the pGreen binary vector series, the plasmid carries origins of replication from pSa and ColE1 to facilitate maintenance in *Agrobacterium* and *E. coli* respectively (Hellens *et al.* 2000b).

Selecting the most appropriate vector for use in plant transformation can be a complex task, and various factors need to be considered. There is no substitute for experience, and factors include host specificity of the bacterial strain, chromosomal background and marker gene, Ti plasmid and marker gene, size of vector plasmid, number of unique restriction sites in the T-DNA, ease of selection of recombinant vectors, bacterial selection marker, the plant selection marker and proximity and orientation to border sequences. An excellent review of binary vectors and their selection for use in plant transformation has been provided recently by Hellens *et al.* (2000a). Compatibility of the vector with the *Agrobacterium* strain(s) that provide optimal T-DNA transfer functions is a critical factor. Obtaining a vector with appropriate selectable marker genes in the T-DNA is a further consideration, although many binary vectors now encompass a wide range of such genes (e.g. pGreen). Beyond these technical considerations the nature of the intended transformation events needs consideration. Is it intended that transgenic lines carry single or multiple (stacked) genes? If the latter is the case, then the choice of vector and selectable

marker can be critical. Stacking multiple transgenes by crossing independent lines is a relatively simple task for many seed propagated crop species such as maize, soybean, oilseed rape and rice, but is not an option for vegetatively propagated plants such as banana and potato. Where the crossing of transgenic lines is intended, it may prove necessary to use vectors that do not contain homologous sequences to avoid problems with gene silencing and therefore have different selectable marker or reporter genes. Similarly, utilizing a vector that allows the use of different selectable markers may also facilitate repeated rounds of transformation provided suitable protocols have been developed to allow efficient use of different marker systems. The need for several protocols, each optimized for a different marker gene, can be avoided if the selectable marker gene can be removed following selection of the transgenic line. Several strategies have been employed to achieve the elimination of marker genes (see Section 4.7).

Significant advances have been achieved in extending the host range of *Agrobacterium* strains through manipulation of bacterial virulence to promote T-DNA transfer from the vector. Some of this has been driven by the need to develop efficient transformation systems for monocots (see earlier). However, some of the same approaches have been applied to dicots and the principal advance – a frequently reported component of transformation protocols – is to induce or elevate the expression of *vir*G which activates the remaining virulence genes (van Wordragen & Dons 1992; Zupan *et al.* 2000). Induction of *vir*G expression, via interaction with VIRA, through the inclusion of acetosyringone in the bacterial inoculation medium can result in elevated rates of transformation. A recent example of the approach has been reported in broccoli (Henzi *et al.* 2000). Expression of *vir*G and *vir*E can be rate-limiting during *Agrobacterium*-mediated transformation where large T-DNA sequences are to be transferred. During transfer of the T-DNA from the bacterium to the plant, the T-strand is coated with VIRE1 protein to protect against nuclease attack. It is estimated that sufficient VIRE1 is produced to coat T-strands of 20 to 25 kb. Consequently, this may impose a limit on the size of T-DNA that can be successfully transferred and integrated into plant genomes. Large sequences of DNA can be successfully transferred from *Agrobacterium* to plant cells using binary bacterial artificial chromosome (BiBAC) vectors. T-DNA regions of 150 kb have been successfully transferred using these plasmids (Hamilton *et al.* 1996). Critical to the successful use of BiBAC vectors is the presence of additional copies of *vir*G or *vir*E on separate replicons that are co-resident in the *Agrobacterium*. As mentioned earlier, Lee *et al.* (2001) recently described a set of intermediate vectors capable of introducing genes into the chromosome of *Agrobacterium*, and demonstrated the potential of manipulating bacterial virulence functions by the stable incorporation of *vir* genes from the Ti plasmid into the bacterial chromosome.

Considerable progress has been accomplished in the design of vectors for T-DNA transfer into plants. It has recently been suggested that the limits of manipulating the bacterial machinery for T-DNA transfer have nearly been reached. Further advances in the efficiency of T-DNA transfer will come from insights into the plant genes involved in the interaction with the bacterium, particularly at the stage of T-DNA

integration into the genome. The mechanisms by which T-DNAs are integrated into the plant genome are poorly understood and little is known of the plant genes that regulate the process. Integration of T-DNA is thought to occur by non-homologous or illegitimate recombination (Gheysen *et al.* 1991; van Attikum *et al.* 2001). Processing of the T-strand can be inaccurate, and T-DNAs with incomplete sequences or sequences from the vector backbone can be incorporated into the plant genome (Smith *et al.* 2001). These events are undesirable for applications of transformation technology, and considerable effort has to be made to detect and eliminate them.

4.3 Direct DNA delivery

Several methods of direct gene delivery have been developed. In general terms, many of these techniques are highly effective methods for transferring DNA into plant cells, but there is quite wide variation in the efficacy with which they can be applied to plant genetic engineering. Systems that produce stable plant transformants are primarily those based around the use of protoplast cultures and particle bombardment into intact tissues.

4.3.1 Particle bombardment

The use of particle bombardment was developed as an alternative means of gene delivery to by-pass limitations imposed by *Agrobacterium*-host specificity. Although many of these limitations have been overcome, gene delivery by particle bombardment has become a widely used technique. Particle bombardment has many applications in plant transformation, including transformation of previously difficult-to-transform monocot and legume species (Christou 1995). The method has found commercial application, being used to produce maize resistant to European corn borer, using *Bacillus thuringiensis* genes encoding insect-specific toxins (either Cry1Ab, Cry1Ac, Cry9C or Cry1F protein), and herbicides, using genes encoding 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) from *A. tumefaciens* strain CP4 and phosphinothricin acetyltransferase (PAT) from *Streptomyces hygroscopicus.*

Microprojectile-mediated gene delivery results when microprojectiles, usually tungsten or gold particles coated with DNA, are propelled at the target tissue by acceleration. Numerous instruments are available for microprojectile-mediated gene delivery, but these devices differ principally in the way that the acceleration force is applied and whether a vacuum chamber is employed. The acceleration force can be provided either by gunpowder, electrical discharge or by gases such as CO_2 or helium. While the method can be applied to a range of tissues, success depends upon optimizing DNA delivery with recovery of transgenic cells. Factors affecting the efficiency of microprojectile-mediated transformation have been reviewed by Sanford *et al.* (1993) and Southgate *et al.* (1995). It is easier to achieve transient expression following microprojectile bombardment, but only a few of these events

become stable transformation events, as few as 1 in 10 000. Stable transformation is frequently restricted by the damage caused to the tissue following bombardment that reduces the viability of the tissue and individual cells receiving DNA. This in turn reduces the subsequent ability of tissues to proliferate and give rise to fertile transgenic plants. Alterations to the physical and biological parameters have had major impacts upon the efficacy of the systems. For example, transformation is frequently increased following pre-culture of the explant material or subjecting the target tissue to osmotic pretreatment either by drying or culturing in the presence of an osmoticum. The use of smaller projectiles (particularly inert gold particles) and adoption of baffling screens have been shown to improve rates of biolistic transformation. Several alternative projectiles have been tested for the delivery of genes to plant cells (Kikkert *et al.* 1999).

Particle bombardment is a simplified physical process of DNA delivery compared to the biological interactions that occur in *Agrobacterium*-mediated transformation. The approach is only limited by the regeneration capacity of the target tissue and the frequency of stable transformation. The tissue culture stage for regeneration of transgenic lines is frequently the time-consuming step in most transformation procedures. Apical meristems are attractive targets for transformation because their use simplifies regeneration procedures. Gene transfer to apical meristems has been achieved using particle bombardment (Sautter *et al.* 1995), and in some species transgenic lines have been recovered (Zhong *et al.* 1996). However, the use of meristems as a target for particle bombardment generally results in the recovery of chimeric transgenic lines; whilst they contain transgenic tissues these do not necessarily contribute to the germline and recovery of stable transformants can be problematic. Strategies have been developed to maximize recovery of transgenic sectors contributing to germline transformation in maize (Lowe *et al.* 1995).

Molecular analysis of plants obtained by particle bombardment generally reveals strikingly different patterns of gene integration compared to those achieved through *Agrobacterium*-mediated transformation. Transformation by particle bombardment often leads to integration at a single locus of complex arrays of multiple copies of the introduced genes including the plasmid vector, often fragmented and rearranged (Pawlowski & Somers 1996; Smith *et al.* 2001). Insertion of multiple copies at a single locus prevents their segregation in subsequent generations and may also result in unstable and unpredictable patterns of gene expression as a consequence of several mechanisms (see Matzke & Matzke 1995). Genetic instability can arise when homologous sequences between the multiple copies recombine. Co-suppression or gene silencing may result when multiple copies of transgenes interact to inactivate themselves or related host genes.

4.3.2 Protoplasts

Systems for the transformation of plant protoplasts are based on the potential ability of single plant cells to regenerate whole plants – that is, totipotency. DNA transfer is facilitated by removal of the cell wall following mechanical and enzymatic deg-

radation, and in general the process generates large cell populations as targets for transformation and regeneration. Protoplast-based systems for transformation are technically very demanding, but advances in the regeneration of both dicots and monocots from protoplasts have been made. Protoplasts are amenable to *Agrobacterium*-mediated transformation but several direct methods for gene transfer to protoplasts have been developed. Protoplasts are usually isolated from established suspension cultures derived from callus and leaf mesophyll cells, although a wide range of tissues has also been used. Direct DNA uptake by protoplast cells is mediated by treatment with polyethylene glycol (PEG), electroporation and liposomes.

Gene transfer into protoplasts using PEG is based on the ability of this chemical to increase the permeability of cell membranes to allow entry of free DNA molecules. PEG itself can be phytotoxic, and an important consideration in optimizing protocols is adjusting PEG concentrations to maximize cell viability and recovery of transgenic cells. In general, optimal concentrations of PEG are in the range of 15 to 25%; high concentrations will decrease cell viability, while low concentrations lead to inefficient gene transfer. Rates of protoplast division and regeneration are greatly influenced by culture media components and these must be optimized. The buffers used to protect protoplasts from cell lysis also exert an influence on the efficiency of gene transfer using PEG, and these too must be adjusted (Zhang & Wu 1988). Electroporation is an alternative means of increasing permeability of cell membranes to induce gene transfer to protoplasts. Short high-voltage electrical pulses applied to protoplasts cause temporary pores to open, allowing entry of DNA into the cell. Optimal conditions for gene transfer using electroporation of protoplasts vary considerably between species, and require optimization of biological and physical parameters. Protoplast size and competent state varies between species, and this in turn alters the requirements for electroporation in terms of adjustments to the duration and strength of the electrical pulse applied. Gene transfer to protoplasts has also been achieved following mild sonication and microinjection.

Initially, gene transfer techniques for protoplasts were established for monocotyledonous species because of problems with *Agrobacterium*-mediated transformation. Techniques have been established in a range of both dicots and monocots that include maize, rice, soybean, brassica and vegetable crop species (Fromm *et al.* 1985; Rasmussen & Rasmussen 1993; Bittencourt *et al.* 1995; Rao *et al.* 1995). Protoplast systems have found wide application in plant biology research laboratories for the functional analysis of genes and gene constructs (Fischer & Hain 1995). The major disadvantage of protoplast systems for plant genetic engineering is that the culture systems are technically demanding and time consuming. Further, as with *Agrobacterium*-mediated systems, efficient regeneration methods have to be established and these vary between genotypes of the same species. Genotypic variation in protoplast regeneration in japonica varieties of rice has been a major restriction to the application of these approaches (Christou 1996). Nevertheless, methods for the regeneration of stable transformed lines have been developed for several species (e.g. Dhir *et al.* 1991; Mukhopadhyay *et al.* 1991; Datta *et al.* 1992; Asano *et al.* 1998; Tiwari *et al.* 2001).

In some plant species protoplast transformation has proven to be highly efficient, and a notable example occurs in sugar beet. Methods of isolating guard cell protoplasts from leaf tissue using a centrifugation step provided access to cells that were highly responsive to regeneration (Hall *et al.* 1997). High transformation frequencies have been obtained following PEG-mediated DNA transformation applied to protoplast populations enriched specifically for stomatal guard cells (Hall *et al.* 1996) and have overcome problems with genotypic variation in *Agrobacterium*-mediated transformation. These observations led Hansen and Wright (1999) to suggest that the method may have application to other recalcitrant species.

4.3.3 Alternative methods for direct gene transfer

Various workers have attempted to produce more efficient or economical means of plant transformation, often in an attempt to overcome restrictions encountered in the application of either *Agrobacterium*- or microprojectile-mediated transformation. Alternative DNA delivery techniques include the use of silicon carbide fibres (also known as whiskers), intact tissue electroporation, electrophoresis and microinjection; the origin and basis of these systems have been reviewed by Songstad *et al.* (1995). The attraction of these systems is that they have the potential to be species- and variety-independent and, in the case of whiskers, can also provide a very simple and rapid means of producing stable transgenic cell lines. However, integration of DNA into the host genome is not a benign process, and there is increasing evidence to suggest that plant genes participate in the regulation of the process (Gelvin 2000; Cogan *et al.* 2001). Consequently, due regard must be paid to the impact that genotypic variation can exert on steps of the process beyond delivery of DNA into the plant cell.

Microinjection has been used to successfully deliver DNA into a range of target cells and tissues. In some cases, DNA delivery is effective but stable integration of genes does not occur (Jones-Villeneuve *et al.* 1995). Nevertheless, successful systems have been established for protoplasts of both dicots and monocots (Crossway *et al.* 1986; Kost *et al.* 1995; Holm *et al.* 2000). Whilst the approach has promise, it is not widely practised as it is especially demanding of both time and technical skills. Microinjection is attractive where tissue culture systems for a particular species or genotype are limiting, especially where cell types that regenerate along natural lines can be identified, as in the case of isolated zygotes (Leduc *et al.* 1996). Indeed, the ability to deliver genes into such tissues provides an attractive means of studying the regulation of zygotic development (Scholten & Kranz 2001).

4.4 Perspectives

Some attempts have been made to combine the best features of direct and indirect methods of transformation. Agrolistics developed by Hansen and Chilton (1996) combines the advantages of *Agrobacterium*-mediated transformation with the high

efficiency of biolistic DNA delivery. In this approach, the virulence genes *vir*D1 and *vir*D2, that are essential for T-strand excision from Ti or Ri plasmids of *Agrobacterium*, are co-delivered to the target material under the control of a constitutive plant promoter. VIRD1 and VIRD2 have been shown to function *in planta* to produce the characteristic strand-specific nicking at the RB sequence, similar to that which leads to formation of the T-strand in *Agrobacterium*. Transformed tobacco calli were obtained after co-delivery of *vir*D1 and *vir*D2 with selectable marker genes flanked by RB and LB sequences. Examination of insertion sites revealed that some exhibited right junctions with the plant genomic DNA that corresponded precisely to the sequence expected for T-DNA insertion events. Such events were termed 'agrolistic' inserts as opposed to biolistic insertion events. Around 20% of the transgenic lines obtained contained only agrolistic insertion events, whilst a further 20% contained both agrolistic and biolistic events. Agrolistic transformation can be used with any target tissue that is susceptible to biolistic transformation. Its principal advantage is the elimination of extraneous vector sequences from the insertion events and a reduction in copy number that will lead to increased stability of expression and inheritance of the transgene. The authors point out that additional copies inserted as biolistic events would likely occur at unlinked loci and could be segregated from agrolistic events after genetic crossing of the transgenic line.

Escudero *et al.* (1995) observed that T-DNA transfer from *Agrobacterium* could occur intracellularly following microinjection of the bacterium into tobacco cells. Intracellular T-DNA transfer from the bacterium was absolutely dependent on *vir* gene induction by pretreatment with acetosyringone at acidic pH. Transfer of T-DNA occurs without binding of the bacterium to the plant cell, suggesting that the approach might have potential in situations where host-cell recognition is not sufficient to induce the natural transfer mechanism from the bacterium. Such problems have been encountered in monocots, particularly in maize where the undifferentiated meristematic cells of early embryos are not competent for transformation by *Agrobacterium*. Microinjection of pre-induced bacterium into early embryo meristematic cells has proven an effective means of obtaining T-DNA transfer (Escudero *et al.* 1996). Interestingly, the approach is genotype-sensitive, and efficient transfer of T-DNAs occurred only in the line A188. The same line was also found to be competent for T-DNA transfer following infiltration of the bacterium. Crossing non-responsive genotypes to A188 does not restore competency for T-DNA transfer, which suggests that the trait is recessive. The assays used did not allow for selection of stable integration events.

4.5 Alternative approaches to plant genetic engineering

4.5.1 In-planta technologies

Of the transformation systems developed to date, most require the use of plant tissue culture (Hansen & Wright 1999) to regenerate whole plants from single transformed

cells. Many successful methods for the regeneration of plant cells to whole plants have been developed. Unfortunately, during regeneration of plant cells various changes can occur, and these have included the regeneration of polyploids or aneuploids, plants with mutations in their DNA, or changes in the pattern of gene expression. Typically, methods require substantial optimization to avoid or minimize these problems in the hands of experienced cell biologists. Regeneration of plant cells is also time-consuming and resource-intensive. The practical knock-on effect of the problems encountered during regeneration is that it is necessary to generate and screen several independent transgenic lines in order to find those that have suffered either no or minimal genetic change (Birch 1997). For the commercial production of transgenic lines, this screening must be extended further to identify lines with simple insertion patterns and freedom from vector backbones. These difficulties serve to limit the efficiency of plant transformation.

A number of laboratories have pursued plant transformation methods that avoid tissue culture or regeneration in an effort to overcome the associated problems. The use of pre-existing meristems or gametes, that are 'programmed' for development into plant structures, as targets for transformation has been investigated by several workers (Zhou *et al.* 1983; Chowrira *et al.* 1995; Escudero *et al.* 1996; Touraev *et al.* 1997; Hu & Wang 1999). Development of whole plants from meristems or gametes generally avoids the problems encountered with regeneration from single cells or tissues. Despite the obvious attraction, these approaches have proven difficult to reproduce and are practised by only a few laboratories. The quest for increases in the efficiency of transformation has become more urgent as genomics programmes begin to deliver large numbers of genes with potential for crop improvement. Whilst transient transformation assays can provide a means of rapidly determining gene function, it is becoming increasingly desirable to produce uniformly transformed plants in which gene function can be determined through development and under a variety of environmental conditions, and particularly under conditions of abiotic and biotic stress.

The first methods of transformation that avoided a tissue culture stage were successfully developed for the model plant *Arabidopsis thaliana*. The various stages in the development of in-planta transformation of *Arabidopsis* have been reviewed by Bent (2000). The first report of in-planta transformation of *Arabidopsis* involved the treatment of seed with *Agrobacterium* (Feldmann & Marks 1987). The method had a relatively low efficiency and was variable between different rounds of transformation, but the approach produced a collection of *Arabidopsis* insertional mutants where the inserted T-DNA disrupted the function of native genes. Subsequently, alternative methods were developed. Chang *et al.* (1994) and Katavic *et al.* (1994) reported successful transformation following the excision of developing inflorescences and application of *Agrobacterium*. Several rounds of inflorescence excision and *Agrobacterium* inoculation were carried out on individual plants that were then allowed to flower and set seed. Although successful, the methods were no more productive than tissue culture-based protocols. The 'quantum leap' for in-planta transformation of *Arabidopsis* came with the discovery that young inflorescences of

Arabidopsis could be infiltrated with *Agrobacterium* under vacuum (Bechtold *et al.* 1993). Plants could then be recovered and allowed to flower and set seed. Transgenic plants could be identified in the next generation, either by germination of seed on antibiotic media or following spraying of plants with herbicide depending upon the selectable marker gene utilized (Bouchez *et al.* 1993). The simplicity of in-planta methodology for transformation of *Arabidopsis* was further extended when it was found that simply dipping or spraying inflorescences with *Agrobacterium* was sufficient to allow recovery of transgenic seed (Clough & Bent 1998). A range of ecotypes and bacterial strains have been used effectively, although notable differences were observed (Clough & Bent 1998) which indicated that the method also suffers from genotypic influences. These methods have proved easy to replicate, and large numbers of transgenic lines were produced without the problems associated with regeneration from single cells; consequently, *Arabidopsis* transformation was revolutionized (Bent 2000).

Transformation events are recovered from in-planta transformation by screening seeds/seedlings of the subsequent generation. The means of selection depends upon the marker included in the T-DNA of the vector, and usually this is either *npt*II or *bar*. Selection of transgenic lines is then achieved by germinating seedlings on kanamycin-containing media, spraying young plants in the glasshouse with kanamycin or with the herbicide 'Basta'. The latter has proven to be a simple and effective approach. Analysis of transgenic lines obtained via seed transformation and vacuum infiltration methods has shown that events derived from a single plant are usually independent and that these primary transformants carry hemizygous T-DNA insertion events (Feldmann & Marks 1987; Bechtold *et al.* 1993). Primary transformants are uniform in that they carry the inserted T-DNA in the cells of all tissues and organs. The inference drawn from these observations is that the transformation event occurs late in floral development at a time when male and female gamete formation is occurring independently. Transformation events occurring before the divergence of the male and female germline cells would be expected to produce some homozyogous T-DNA insertion events. Furthermore, transformation events must occur at a time before the cell division commences in the fertilized embryo and independent lineages of cells are formed, as transformation at these later stages would result in chimeric plants. Several research groups have set out to determine the target of transformation following floral dipping or vacuum infiltration of *Arabidopsis* flowers. Reciprocal crosses made following transformation whereby the treated plant is used as both pollen donor and recipient to non-transformed plants have revealed much about the site of transformation. Using Agro-inoculated plants as a pollen donor fails to yield transgenic lines, whereas inoculated plants that receive non-transgenic pollen produce transgenic lines (Ye *et al.* 1999; Bechtold *et al.* 2000). These observations indicate that the male germline cells are not the target of transformation. A series of studies utilizing promoters active in floral tissues driving the *gus* reporter gene has shown that the site of transformation is the ovule (Ye *et al.* 1999; Desfeux *et al.* 2000) and that β-glucuronidase (GUS) staining is absent from anthers and pollen. Further evidence using genetic linkage analysis of a marked chromosome revealed that, in

the majority of transformants studied (25/26), inserted T-DNAs are carried on the maternally inherited chromosome. In one of the events investigated, the T-DNA was apparently associated with the paternal chromosome set, suggesting that pollen transformation is a possibility or that integration of T-DNA in this line occurred at a later stage into the diploid genome within the fertilized embryo.

Staining Agro-inoculated *Arabidopsis* plants for GUS activity has also revealed the importance of timing inoculation for efficient transformation. In most plants a large proportion of developing siliques bear no transgenic embryos. Only a few siliques will contain transgenic embryos; between one and seven per silique have been observed (Desfeux *et al.* 2000), and these siliques are distributed unevenly around the plant (Clough & Bent 1998). Particular flowers are evidently more receptive to transformation than others, and it seems this may be related to the timing of inoculation relative to the development of the ovule. On a practical level, repeated applications of *Agrobacterium* at 5- to 6-day intervals increase the frequency of transformation, presumably by targeting more flowers at the right stage of development; however, frequent inoculations are detrimental to the plant (Clough & Bent 1998). In *Arabidopsis*, the developing ovule remains open and thus accessible to any *Agrobacterium* through much of the floral development. At about 3 days prior to anthesis the developing ovary becomes enclosed by the formation of the stigma. Typically, flowers 4 days from anthesis fail to bear transformants, whilst flowers at 5 to 10 days from anthesis do – presumably because the developing ovules remain accessible to invading *Agrobacterium* (Desfeux *et al.* 2000). For floral dip transformation, the optimal stage was reached after removal of the primary bolt and when the secondary bolts were 2 to 10 cm in length and bearing only a few open flowers (Clough & Bent 1998). The floral dip method uses a simplified medium for resuspension of the Agrobacteria, typically a 5% sucrose solution containing 0.05% Silwet L-77, a non-phytotoxic wetting agent. Pre-induction of the *vir* genes appears not to be necessary in this medium (Clough & Bent 1998). Transformation frequencies between 0.5 and 3% are obtained by floral dip transformation of ecotypes such as Col-0, Ws-0 and Nd-0. Interestingly, much lower frequencies of transformation were obtained with Ler-0, and this has been related to the fact that floral development in this ecotype differs from Col-0 (Clough & Bent 1998). The method also seems tolerant of different bacterial strains. GV3101, LBA4404, EHA105 and Chry105 have all been reported to successfully transform *Arabidopsis* after floral dipping.

Arabidopsis in-planta transformation methods are simple and quick, and require only a supply of plants for inoculation. The principal advantages are the ability to avoid a tissue culture stage, ease of selection, and minimal labour and expertise. Such simplicity and ease of operation has proven an attractive stimulus for examining the potential of in-planta transformation in other crop plants. Amongst the potential candidates are rice, wheat, brassica (oilseeds) and soybean. Attempts at floral dipping, vacuum infiltration and related methods have undoubtedly been attempted in many laboratories. An early report described pipetting *Agrobacterium* onto the floral spikelets of wheat where transformation frequencies of 1 to 2.6% were observed (Hess *et al.* 1990), and interestingly the target of transformation was postulated to be

pollen rather than ovules. Successful application of infiltration methods developed for *Arabidopsis* to Pakchoi (Qing *et al.* 2000) raised the prospect of the method being applicable to other members of the Brassicaceae, although so far no successful reports have been formally published. A further stimulus has been provided by the report that the legume *Medicago truncatula* could be transformed following vacuum infiltration of *Agrobacterium* (Trieu *et al.* 2000). The target of transformation appeared to differ from *Arabidopsis* in that homozygous transgenic lines were recovered and the majority of lines were derived from the same transformation event. It has been suggested that transformation may occur at an earlier stage of floral development, possibly before cell lineages that give rise to the female and male gametes diverge, in *M. tunculata* compared to *Arabidopsis* (Bent 2000). Determining the target tissue of transformation will allow for comparison between methods and provide important insights for the development of in-planta techniques in other crops.

The ease with which in-planta transformation methods are developed in other crops may well hinge upon studies aimed at determining stages of floral development and sites within the floral tissues that are competent for *Agrobacterium*-mediated transformation. Bent (2000) proposed the idea of screening floral tissues and specifically monitoring transformation of ovules or progenitor tissues for competence for transformation as a starting point to the evaluation of different methodologies. It is clear that the identification of the target of transformation in *Arabidopsis* provided a means of increasing transformation efficiency. The use of mutants such as CRABS-CLAW in which the female floral tissues are more accessible to the *Agrobacterium* enhances overall efficiency (Desfeux *et al.* 2000). Bacterial strain effects on transformation rates have been minimal for in-planta transformation of *Arabidopsis.* Whilst genotypic differences have been encountered, in-planta transformation has provided a means of transformation in genotypes untransformable by tissue culture methods (Nam *et al.* 1999). It is not clear whether these factors will be important in the development of in-planta methodology in other crop plants. As with all transformation methods, a better understanding of the mechanism of T-DNA integration into the host genome and the basis of genotypic variation for transformation competence will provide further insights into the refinement of approaches to gene delivery.

4.5.2 Plastids

Many characteristics of the plastid genome (plastome) make it an attractive target for insertion of transgenes and several systems have been developed for its transformation. Identical copies of the plastome are found in all cells and all types of plastid (proplastids, etioplasts, amyloplasts, chromoplasts and chloroplasts). Leaf cells may contain as many as 100 chloroplasts each harbouring around 100 copies of the plastome resulting in very high ploidy levels (Bendich 1987; Staub *et al.* 2000). Chloroplast transformation mediated by *Agrobacterium* was first established by Deblock *et al.* (1985), and subsequently stable chloroplast transformation has been achieved following particle bombardment (Sidorov *et al.* 1999; Staub *et al.* 2000) and poly-

ethylene glycol treatment (Koop *et al.* 1996) of tissues and protoplasts respectively. Reliable methods have been developed for tobacco and recently extended to tomato. There are also reports of plastid transformation in potato and *Arabidopsis*, although evidence for transmission of traits to progeny is lacking in these species.

Initial attempts at plastid transformation utilized a selection system based on a mutation in the plastid 16S rRNA that confers resistance to spectinomycin and typically yielded a single transformation event from 100 bombardments (Svab *et al.* 1990). Subsequently, a chimeric bacterial *aadA* gene that encodes aminoglycoside 3'-adenylyltransferase has been used (Svab & Maliga 1993). The *aadA* gene confers spectinomycin and streptomycin resistance to bacteria. Expression driven by the plastid-derived P_{rrn} promoter, from the plastid rRNA operon, resulted in a 100-fold increase in plastid transformation.

The efficiency of the integration process is dependent on the frequency of homologous recombination. Several different flanking sequences have been used to effect homologous recombination of transgenes into the plastome. The pZS197 vector employed by Svab and Maliga (1993) used the *rbcL* gene and *ORF512* as flanking sequences. Transgenic plastomes carrying the inserted fragment were recovered in tobacco following particle bombardment. Additional recombination events were detected between homologous sequences in the *aadA* gene and the native plastid *psbA* gene. These events led to the deletion of genes required for plastid maintenance and could not be maintained in a homoplasmic state. Interestingly, these authors also noted that recombination events between homologous sequences of the promoter regions used were not detected, suggesting that recombination between short sequences of homology does not occur within the plastome. Recently, Ruf *et al.* (2001) have described the use of the *rps14/trnfM* region of the tobacco plastid genome as flanking sequences into which the selectable spectinomycin resistance marker gene (*aadA*) and the polylinker from pBluescript were inserted. The *rps14/trnfM* region is highly conserved amongst plastomes of dicotyledonous plants and this construct was shown to function in tobacco and tomato with high efficiency. The region is less useful for graminaceous species, such as rice and corn, which have non-homologous sequences in this region (Ruf *et al.* 2001).

Initially, transformation of the plastome results in the generation of a heteroplastomic state in the target explant. Successful transformation is dependent on the ability selectively to amplify the transplastomic genome in order to achieve a homoplastomic state. This has been a major obstacle in extending the technology beyond tobacco, and Ruf *et al.* (2001) consider this to be a consequence of limitations in the currently available tissue culture and regeneration systems. The critical factors in extending techniques to tomato were described as the use of:

- extreme low light conditions during selection (25 μE compared with 70–100 μE used for tobacco);
- the extension of the selection phase from 3–5 weeks to 3–4 months; and
- the use of significantly smaller leaf explants (Ruf *et al.* 2001).

Attainment of the homoplastomic state requires molecular analysis of transgenic lines to verify the condition in cultured tissues prior to regeneration. Although this represents a time-consuming step, recent reports indicate that efficiency may be improved through the use of fluorescence-based markers that allow tracking of the transformed and wild-type plastomes (Sidorov *et al.* 1999). Attainment of homoplasty is essential if genetic drift is to be avoided when selection pressure is removed.

Whilst the range of crop plants successfully engineered to contain transgenic plastids is limited, there are several advantages to the approach. Expression of proteins in plastids presents fewer problems compared with nuclear transformation methods. Targeted integration into the plastome, coupled with the lack of a compact chromatin structure, ensure that positional effects do not influence transgene expression (Staub & Maliga 1992). Furthermore, silencing of transgenes inserted into the plastome has not been encountered and transgene expression is stable in progeny (Sidorov *et al.* 1999). In tobacco, polycistronic mRNAs transcribed from genes inserted into the plastome are efficiently translated, offering the prospect of expressing multiple transgenes stacked together as single operons (Staub & Maliga 1995). Legitimate post-translational processing of proteins produced from plastid-located transgenes has been reported (Staub *et al.* 2000), indicating the potential of plastid transformation for the production of high-value pharmaceutical proteins. In addition to these characteristics, plastid transformation permits a higher level of containment of transgenes because, in most higher plants, plastids are inherited from the maternal parent (Daniell & Varma 1998) so that there is no pollen transmission for transgenes.

Insertion of transgenes into the plastome can result in exceptionally high levels of gene expression, and transgene-encoded protein concentrations greater than 40% of the total soluble cellular fraction have been reported (McBride *et al.* 1995; Staub *et al.* 2000). Consequently, plastid transformation is an attractive target for molecular pharming for the production of high-value therapeutic or industrial proteins (Ruf *et al.* 2001). High rates of gene expression in the plastome also have implications for the effectiveness of resistance genes (McBride *et al.* 1995), as increased production of insecticidal proteins in leaves would mean that leaf-feeding insect pests would need to consume less biomass for a given insecticidal dose.

High levels of transgene expression in the plastome are not universally required, however. For many traits, high expression is much less important than ensuring that transgenes are active at specific stages of development. A potential drawback of plastid transformation is the lack of regulatory sequences permitting tissue specific control of expression compared to nuclear transformation. Regulation of gene expression in plastids predominantly occurs at translation. McBride *et al.* (1994) proposed a system whereby silent plastid genes could be switched on by nuclear encoded trans-activation signals targeted to the plastome. The utility of the approach has been demonstrated by the use of a plastid-targeted T7 RNA polymerase and the phage T7 gene 10 promoter fused to the β-glucuronidase reporter gene.

4.6 Selection of transformation events

Transformation events are relatively rare, and it is necessary to have a means of selection to ensure survival of transformed cells and prevent regeneration from non-transformed cells. In the majority of transformation systems developed (see later for exceptions) this has been achieved by linking the transgene of interest to a gene encoding a selectable marker. Many selectable markers have been developed, and considerable effort has been expended on increasing transformation efficiency by optimizing the selection process (for a recent review, see Joersbo 2001). Generally, selection systems employ dominant selectable marker genes that permit the survival of transformed cells either by the enzymatic degradation of toxic selection agents (commonly antibiotics and herbicides) or their modification to much less toxic forms (see Table 4.1). To date the most widely used selectable marker gene has been neomycin 3'-*O*-phosphotransferase, conferring resistance to aminoglycoside antibiotics that include kanamycin, neomycin and G-418 (Fraley *et al.* 1983; Herrera-Estrella *et al.* 1983).

New marker systems for the selection of transformed cells are continually being sought, and these have provided greater flexibility in the development of transformation protocols for previously untransformed plant genotypes. There is also growing interest in delivering multiple transgenes into crop plants, which may require repeat transformation and hence several different selectable markers.

Table 4.1 Dominant selectable marker genes employed in plant transformation systems for the selection of plant cells. Genes confer resistance to antibiotics, herbicides or metabolic inhibitors.

Marker gene		Selection agent(s)
*npt*II/*neo*	Neomycin phosphotransferase	Kanamycin, neomycin, geneticin (G418)
cat	Chloramphenicol acetyltransferase	Choramphenicol
hpt	Hygromycin phosphotransferase	Hygromycin
bla	β-Lactamase	Penicillin, ampicillin
aadA	Aminoglycoside-3'-adenyltransferase	Streptomycin, spectinomycin
dhfr	Dihydrofolate reductase	Methotrexate
bar/pat	Phosphinothricin acetyltransferase	Bialaphos, phosphinothricin, glufosinate
epsps/aroA	5-Enolpyruvate shikimate-3-phosphate	Glyphosate
bxn	Bromoxynil nitrilase	Bromoxynil
als		Sulphonylureas, imidazolinones, thiazolopyrimidines
TDC	Tryptophan decarboxylase	
xylA	Xylose isomerase	
ManA/Pmi	Phosphomannose isomerase	Mannose-6-phosphate
BADH	Betaine aldehyde dehydrogenase	Betaine aldehyde

Several of the more recently developed selectable markers focus on components of plant tissue culture media. Plant cells are auxotropic in culture and require an exogenous supply of several substances, including nutrients, vitamins, growth regulators and a carbon source. The basis of selection is to complement transformed cells with a gene that enables them to produce an essential substance of which they are deprived in the tissue culture medium (Joersbo 2001), and this has been described as positive selection (Haldrup *et al.* 1998). Non-transformed cells are effectively 'starved' of an essential component of growth rather than directly killed by the selection agent. For example, the sugars xylose and mannose have been developed as positive selective agents. Many plant species are unable to metabolize D-xylose, and when their cells are cultured on media containing this as a sole carbon source they do not grow. In contrast, transgenic cells expressing a xylose isomerase gene (D-xylose ketol-isomerase) are able to utilize the keto-isomer of xylose, D-xylulose. Haldrup *et al.* (1998) have used this gene as a selection agent during *Agrobacterium*-mediated transformation of potato, tomato and tobacco. Efficiencies of transformation were greater in potato and tomato compared with the use of kanamycin selection. However, the use of xylose required adjustment of the culture medium for potato, partly to alleviate a toxic effect of xylose.

Mannose itself has no direct adverse effect on plant cells in culture. When added to plant cell cultures it is converted to mannose-6-phosphate, which is not metabolized further. However, accumulation of mannose-6-phosphate in cultured plant cells leads to the cessation of growth (see Reed *et al.* 2001). Mannose-6-phosphate isomerase converts mannose-6-phosphate to fructose-6-phosphate, which is readily metabolized in glycolysis. The gene encoding mannose-6-phosphate isomerase has been successfully used as a selectable marker following the transformation of several species, including maize, wheat, barley, sugar beet, rape seed, potato and *Arabidopsis* (see Joersbo 2001; Reed *et al.* 2001). Although this system is not universally applicable (since mannose-6-phosphate isomerase activity is found in some plant species), increases in selection efficiencies have been reported in several species using either *Agrobacterium*-mediated or biolistic transformation (Reed *et al.* 2001). As with xylose, optimal selection pressure may require combinations of mannose with supplements of other carbon sources because of phytotoxic effects (Reed *et al.* 2001). Preliminary safety assessments indicate that the enzyme has no adverse effects with respect to mammalian toxicity or allergenicity. Furthermore, no unintended effects on the agronomic or nutritional composition of transgenic plants produced using mannose selection are found (Reed *et al.* 2001).

Recently, a further alternative to the antibiotic and herbicide resistance markers has been developed based on the application of 2-deoxyglucose (2-DOG) to plant culture media. Uptake of 2-DOG leads to its phosphorylation to 2-DOG-6-phosphate by hexokinase in the plant cytosol. 2-DOG-6-phosphate is a metabolic competitor of glucose-6-phosphate and is known to severely impair plant growth due to multiple effects in metabolism. A gene (*DOG*^{R}1) encoding 2-DOG-6-phosphate phosphatase originating from baker's yeast has been used to confer 2-DOG resistance in transgenic plants of tobacco and potato (Kunze *et al.* 2001), and selection was applied at

specific stages of transgenic plant production. Comparable rates of transformation to kanamycin selection have been claimed, although 2-DOG sensitivity varies between different genotypes. Initial toxicological data and the origin of the gene are favourable attributes of 2-DOG selection for plant transformation.

Some selection systems have been based on the requirement for cytokinin activity in cultured cells for the production of shoots. For example, the isopentyl transferase (*ipt*) gene from the Ti plasmid of *Agrobacterium tumefaciens* has been placed under the control of an inducible promoter (the dexamethasone (DEX)-inducible system; Aoyama & Chua 1997). Induction of the *ipt* gene by DEX addition leads to increased shoot formation of transformed cells without the need for exogenous additions of cytokinin to culture media (Kunkel *et al.* 1999). The use of inducible expression of *ipt* overcomes earlier problems encountered when constitutive expression led to an abnormal phenotype and altered fertility of the transgenic plants (Smigocki & Owens 1988). Another cytokinin-based approach has been developed by Joersbo and Okkels (1996). Here, a glucuronic acid-cytokinin complex is supplied to the culture medium. The glucuronide group is readily hydrolysed by the *gus* reporter gene encoding β-glucuronidase (Jefferson *et al.* 1987), and release of active cytokinin leads to regeneration and growth of transgenic cells. The approach has the advantage of utilizing the *gus* gene as both a screenable and selectable marker and so reduces the number of genes transferred to the transgenic plant (Joersbo 2001). Although an attractive system, the selection agent is expensive and is not yet commercially available. For selection based on cytokinin action, escapes can arise where neighbouring transgenic cells either supply sufficient cytokinin to meet the needs of regeneration in non-transgenic cells of the explant or form complexes with them during regeneration (Berthomieu *et al.* 1994).

4.7 Marker-free transformation

The use of a marker gene to identify transgenic plant cells is an essential component of the transformation process. Of the economically important crops for which transgenic lines have been commercialized, most retain the selectable marker gene used within the laboratory phase of the transformation process. The use of antibiotic- and herbicide-selectable marker genes has prompted a wide-ranging debate over their biosafety. There is a considerable body of evidence to demonstrate the safety of these genes (Flavell *et al.* 1992; Fuchs *et al.* 1993). Nevertheless, a desire to remove antibiotic markers from GM products has been clearly expressed (Royal Society 1998). As well as the perceived biosafety issue, a second driving force for their removal is a need to develop protocols that allow the repeated transformation of the same genotype in order to stack transgenes. In conventional transformation strategies this approach would result in the simultaneous stacking of homologous marker gene sequences. A second use of the same marker is clearly not possible for selection in culture, and repeat use would inevitably risk homology-dependent transgene silencing (Matzke & Matzke 1995). Several approaches to avoid these problems have been

suggested by Yoder and Goldsborough (1994). Ebinuma *et al.* (2001) have recently reviewed these strategies in the light of advances in the development of positive markers that are an alternative to antibiotic and herbicide resistance marker genes. Systems for the removal of selectable marker genes include the use of co-transformation, site-specific recombination, transposable elements and intra-chromosomal recombination. In many cases specific vectors have been engineered to facilitate the application of these solutions.

4.7.1 Co-transformation

The simplest approach to generating plants free of marker genes is co-transformation (sometimes also referred to as 'two T-DNA' transformation). Here, the selectable marker gene and the other gene(s) of interest are simultaneously delivered into the plant cell on separate T-DNAs and integrate into the genome. If the T-DNAs integrate at loci that are genetically unlinked, then during subsequent sexual crossing (or selfing) the two T-DNAs will segregate from each other. In this way progeny that lack the selectable marker but are transgenic for the gene of interest can be identified in the second generation. The method has been used successfully with both *Agrobacterium*-mediated transformation and direct DNA delivery, particularly by microparticle bombardment, although the efficiency of the approach varies between the two DNA delivery systems.

Several strategies have been developed for *Agrobacterium*-mediated co-transformation, and these differ in the detail of the bacterial strains and vector plasmids used. Co-transformation can be achieved by co-inoculating explants with two bacterial strains each carrying a vector. In this situation, the T-DNA of one vector carries the selectable marker and that of the other the gene of interest. This strategy has been used for tobacco (Depicker *et al.* 1985; McKnight *et al.* 1987), oilseed rape (Deblock & Debrouwer 1991), *Arabidopsis* (DeNeve *et al.* 1997) and rice (Komari *et al.* 1996). These studies have shown that a single plant cell can be transformed with T-DNAs from two separate *Agrobacterium* strains at frequencies ranging from 2% to 85%. Segregation of T-DNAs has been demonstrated in some cases, indicating that integration occurs at separate loci (McKnight *et al.* 1987; DeNeve *et al.* 1997), although Deblock and Debrouwer (1991) have shown that in oilseed rape only 23% of the integrations were at independent loci.

Several authors have reported higher frequencies of co-transformation than would be expected on the basis of independent transformation events (Depicker *et al.* 1985; Deblock & Debrouwer 1991) and it has been suggested that the frequency of integration at unlinked sites is influenced by the choice of bacterial strain. More than double the number of co-transformed lines produced by octopine strains (Komari *et al.* 1996; Daley *et al.* 1998) produced progeny in which segregation of the T-DNAs occurred compared with data for nopaline strains (Deblock *et al.* 1989).

Co-transformation may also be achieved using a single bacterial strain containing two separate T-DNA-bearing plasmids. Wild-type strains of *A. tumefaciens* and *A.*

rhizogenes will transfer T-DNAs from both their virulence plasmids (Ti or Ri) and from binary vectors *in trans* (DeFramond *et al.* 1986; Puddephat *et al.* 2001). Segregation of T-DNA from the virulence and the binary plasmids has been demonstrated in tobacco and brassica (Boulter *et al.* 1990; Puddephat *et al.* 2001). The frequency of co-transformation in the single wild-type strain approach is reported to vary from 63% to 95%. Disarmed strains of *A. tumefaciens* carrying two binary vectors have also been used. Daley *et al.* (1998) co-transformed oilseed rape and tobacco explants with a single strain carrying two 'binary' vectors – a strategy made possible by designing the binary vectors with different origins of replication to ensure maintenance in the bacterium. Frequencies of co-transformation, following kanamycin selection, were 62% for oilseed rape and 52% for tobacco. Some 40% and 58% of the co-transformed lines (respectively) produced progeny that contained only one of the T-DNAs, indicating that integration had occurred at independent loci.

Several wild-type *Agrobacterium* strains possess virulence plasmids that naturally carry two T-DNA regions and the capacity to transfer both regions to plant cells. Several binary vectors have also been constructed that contain multiple T-DNA regions and exploit the natural capacity of a single bacterium to transfer multiple T-DNAs of different origins. In essence, the strategy is based around one T-DNA region carrying the selectable marker gene while the remaining region(s) carries gene(s) of interest. The approach has yielded relatively high rates of co-transformation with Depicker *et al.* (1985) reporting co-transformation frequencies in tobacco of 67% to 73% and Komari *et al.* (1996) reporting rates of 50% and 47% in tobacco and rice respectively. Segregation of the co-transformed T-DNAs has been demonstrated indicating integration occurs at unlinked loci. Komari *et al.* (1996) reported that, of the co-transformed lines tested, between 56% and 100% produced progeny expressing genes carried on just one of the T-DNAs.

In a series of elegant experiments, McCormac *et al.* (2001) examined the relationship between the relative sizes of the T-DNA regions used in co-transformation using the single bacterium and single vector strategy. The results demonstrated that recovery of co-transformed events is favoured when the T-DNA region carrying the selectable marker gene is larger than the region carrying the gene of interest, data that are consistent with observations of co-transformation frequencies using wild-type strains of *A. rhizogenes* (Puddephat *et al.* 2001). Furthermore, this strategy also favours recovery of co-transformed lines in which the T-DNAs are integrated at independent loci and consequently facilitates genetic segregation in the subsequent generation (McCormac *et al.* 2001).

Frequencies of co-transformation vary widely (from 10 to 95%) for the methods described above, but all demonstrate the possibility of losing T-DNA carrying selectable markers by simple segregation if transgenic plants are taken through a sexual generation. Co-transformation adds a layer of complexity to the development of transformation protocols and it is also only applicable to crop plants that can be sexually propagated and is most readily achieved in those that have relatively short life cycles.

4.7.2 Site-specific recombination

This strategy is based on the use of the two-component *cre*/*lox* site-specific recombination system of the bacteriophage P1 that consists of a recombinase (CRE) and its DNA recognition sites (*loxP*). DNA sequences between two adjacent *loxP* sites can be excised by the action of CRE. Consequently, marker gene removal is facilitated by flanking a selectable marker with two *loxP* sites in a T-DNA region also containing the gene of interest. This construct is used to generate the first transgenic line. A second transformation event is then required to introduce the *cre* gene. This may be achieved by re-transforming the initial line or through the generation of an independent line carrying the *cre* gene and subsequent crossing. Following removal of the *loxP*-flanked marker gene, a further generation is required to segregate the gene of interest from the *cre* gene and marker-free plants can then be identified in the progeny. The effectiveness of CRE to mediate excision of *loxP*-flanked marker genes has been successfully demonstrated in tobacco and *Arabidopsis* (Dale & Ow 1991; Russell *et al.* 1992b). When the *cre* gene is introduced by re-transformation, the excision of the *loxP*-flanked marker gene occurs early in the regeneration of shoots and is relatively efficient (90–95%). However, recovery of marker-free plants following crossing with lines containing *cre* has been less efficient. Russell *et al.* (1992b) obtained chimeric lines of tobacco that contained both the complete T-DNA sequences, including the *loxP*-flanked marker gene, and the excision product. The efficiency of excision in the crossing strategy is related to the function of the *cre* line, which appears to be a function of the expression of the *cre* gene in different cells and during different developmental pathways (Russell *et al.* 1992b). Experience to date indicates that excision occurs more efficiently in re-transformed tissues *in vitro*, suggesting that the processes may occur more readily in callus than organized structures such as meristems and embryos (Ebinuma *et al.* 2001). Further studies with tobacco have also shown that marker-free plants can be recovered using the Cre/*lox* system following a re-transformation strategy in which transient expression of *cre* leads to excision of the *loxP*-flanked marker gene (Gleave *et al.* 1999).

More recently, Gilbertson *et al.* (2002) presented data that demonstrated that the Cre/*loxP* system was much more efficient following transformation by *Agrobacterium* than particle bombardment. It was presumed that the simpler insertion events following *Agrobacterium*-mediated transformation favoured successful recombination. Furthermore, Cre/*loxP* recombination events to remove marker genes were almost 100% efficient in monocots such as corn and wheat, but much less efficient in dicot crop plants. In dicots, partial excision events were detected when *cre* expression was driven by the *nos* promoter; the use of a 'strong' promoter such as CaMV 35*S* to drive *cre* expression was more successful, confirming the earlier observations of Russell *et al.* (1992b).

In addition to the crossing and re-transformation strategies, inducible expression systems have been developed for *cre*. These autoexcision strategies are dependent on the efficiency with which *cre* expression is induced. Chemical- and heat shock-induced promoters have been used successfully (Gilbertson *et al.* 2002; Wang *et al.*

2002). Similar approaches have been tried previously with alternative recombination systems; for example, a Multi-Auto-Transformation (MAT) vector system combines the use of oncogenic genes from either *A. tumefaciens* or *A. rhizogenes* for selection of transformation events based on altered morphology with chemically induced R/RS recombination (Ebinuma & Komamine 2001). The MAT vectors centred on the use of the *ipt* gene have also led to improved generation of transformation events, but the strategy offers wider flexibility and could also be combined with other selectable markers. Chemical induction of the recombinase activity leads to the excision of the marker and recombinase producing marker-free plants or cell lines.

4.7.3 Transposable elements

Yoder and Goldsbrough (1994) have reported the development of a marker-free transformation strategy built around the *Ac/Ds* family of maize transposable elements. In this system, the *Ac/Ds* functions are used to drive the excision of either a marker gene or a gene of interest flanked by *Ds* elements in the presence of the *Ac* element. Following excision, the *Ds*-flanked regions may be transposed to other genomic locations. Where transposition results in integration of the flanked sequences at unlinked loci the marker gene and gene-of-interest may segregate from each other in the subsequent generation. The system has been used successfully in tomato (Goldsbrough *et al.* 1993) and marker-free plants recovered. In order to obtain stable transformants, this approach requires that the selected segregants do not contain copies of both Ac and Ds since this could lead to further transposition events.

4.7.4 Perspectives

Several approaches have been developed for the removal of marker genes following plant transformation. Co-transformation probably offers the simplest route to achieving marker-free plants in crops that are seed-propagated. As with the methods based on transposable elements, co-transformation requires an additional generation to generate the marker-free lines. An attraction of the recombination strategies is that they can be used within the tissue culture phase to generate the marker-free events. However, questions remain as to the fate of the excised fragment of the DNA bounded by the recombination elements. Are these fully or partially degraded? Can any of the fragments re-insert into the genome? These questions will need to be answered within the regulatory framework for commercial GM crop plants before such material is released commercially.

4.8 Prospects for improving the efficiency of transformation

So where are we with plant transformation? Techniques have existed for the delivery of transgenes into all the major crop plants (at least for some genotypes) since the mid-1990s. The technology to remove marker genes associated with the process of

transformation is becoming more widely used within the technologies established for crop plants. Simpler methods of transformation hold great promise for the expansion of gene testing following genomics programmes, particularly from *Arabidopsis* and rice. A major aim for the future will be to achieve targeted integration of transgenes either to specific loci or to effect gene replacement within the genome. Our ability to achieve these goals is still limited, but active research programmes are beginning to realize these aims by enhancing our understanding of the mechanisms involved. What is clear is that our current understanding of the mechanisms behind transgene integration is incomplete and that as we understand more new technologies will emerge.

We have gained significant molecular and genetic insights into the natural mechanisms by which *Agrobacterium* transfers DNA into plant cells. Knowledge of the action of the virulence genes and mechanism of T-DNA transfer has been instrumental in the improvement and extension of gene transfer to many species, including monocots (as discussed above). Interestingly, it has also found specific application in direct gene transfer methods (Hansen & Chilton 1996). With this knowledge, we are able to transform all the major crop plants, yet the efficiency of transformation remains variable between different genotypes of the same crop species. For *Agrobacterium*-mediated transformation it has been suggested that we are reaching the limit of potential gains to be made from an understanding of the bacterial genes involved and that further increases in transformation efficiencies are most likely to be achieved from a greater understanding of the processes by which transgenes become incorporated into the plant genome. Gelvin (2000) has recently provided an excellent account of the roles of *Agrobacterium*-encoded proteins in gene transfer and reviewed the roles of plant genes in the regulation of the process.

4.8.1 Genotypic variation for plant transformation

The concept of genotypic variation in the efficiency of gene transfer mediated by *Agrobacterium* has been discussed in many accounts of the development of transformation systems. Species differences in the susceptibility to infection by either *A. tumefaciens* or *A. rhizogenes* are well documented (Anderson & Moore 1979; Tepfer 1990; Porter 1991) and are particularly notable between dicot and monocot species. Genotypic variation for tumorigenesis or hairy root induction also occurs within a species, as different cultivars or ecotypes can exhibit vastly different rates of infection. For example, Byrne *et al.* (1987) found striking differences in the frequency of tumorigenesis between cultivars of soybean (*Glycine max*) inoculated with *A. tumefaciens*, while Ohlsson and Eriksson (1988) reported similar differences between cultivars of *Sinapis alba*, *Brassica campestris*, *B. napus* and *B. oleracea* where tumorigenesis ranged from 0% to 100% infection. Such results are common to a number of species that also include woody plants (Bergmann & Stomp 1992; Beneddra *et al.* 1996), monocots (Li *et al.* 1992), legumes (Hood *et al.* 1987) and the model species *Arabidopsis* (Nam *et al.* 1997). Similar experiences have been

encountered with the induction of hairy roots following infection by *A. rhizogenes*, as was the case in pumpkin (*Cucurbita pepo*) (Katavicacute & Jelaska 1991).

The efficiency of plant transformation protocols is usually determined by the overlap between cells that are competent to receive transferred DNA and those that are competent to regenerate whole plants. Several authors have examined the competence of cells for DNA uptake (Colby *et al.* 1991; Davis *et al.* 1991; Sangwan *et al.* 1992; Dekathen & Jacobsen 1995). Whilst there are pronounced differences between plant species and even genotypes (e.g. as in brassica; Puddephat *et al.* 2001), stimulation of cell division by phytohormones and/or wounding often leads to increased transformation efficiencies (Chateau *et al.* 2000). The implication is that integration of transferred DNAs may be influenced by a specific stage of the cell cycle. Villemont *et al.* (1997) have studied the influence of the cell cycle on T-DNA transfer and integration in synchronized cell suspensions of petunia. Cell cycle-specific inhibitors were used to block cells at specific stages. Cells blocked at late G_1 could not express T-DNA genes and were not stably transformed, whereas cells blocked at M phase could not be stably transformed but did express T-DNA genes transiently. In contrast, cells released to S-phase were stably transformed. The results suggest that T-DNA integration requires cell division (M phase) and transient expression requires DNA synthesis (S phase). However, it is not yet clear what mechanisms are operating behind these observations, particularly with respect to the fate of single- or double-stranded T-DNAs. Clearly, environmental, physiological and cellular effects on transformation will complicate studies aimed at identifying and determining the role of plant genes in the process. Nevertheless, in several species a genetic basis to transformation efficiency has been firmly established including pea (Robbs *et al.* 1987), soybean (Mauro *et al.* 1995a, b), brassica (Cogan *et al.* 2001) and *Arabidopsis* (Nam *et al.* 1997).

Using a genetic approach for the analysis of plant factors controlling transformation efficiency, Cogan *et al.* (2001, 2002) have established a quantitative basis to transgenic root induction mediated by *A. rhizogenes* in double haploid (DH) lines of *Brassica oleracea* and identified a series of quantitative trait loci (QTL) controlling the process. Individual DH lines with high efficiencies for stable transformation (>85% of inoculated explants) were identified, even from parental lines otherwise recalcitrant to transformation. Using a DH mapping population, these authors identified QTL controlling the efficiency of transgenic root induction on seedling explants mediated by *A. rhizogenes.* Transgenic root induction was then followed in a set of mapping lines with differential recombination events and chromosome substitution lines. Two of the QTL identified were found to be responsible for increasing the efficiency of T-DNA transfer as measured by the number of transient transformation events using a GUS reporter gene. A further QTL was found to increase the efficiency with which transient transformation events were converted to stable events. Substitution events in the region of this QTL abolished stable transformation. The use of substitution lines allows finer mapping of the QTL and with it the prospect of exploiting the synteny between *Brassica* and *Arabidopsis* to identify the plant genes involved.

The genotypic variation encountered within transformation protocols, particularly *Agrobacterium*-mediated transformation, raises some interesting points with respect to future improvements of the technology. Gene delivery is not usually the rate-limiting step in plant transformation. In both direct and indirect transformation methodologies the transfer and transient expression of gene sequences tends to be high, but recovery of stably transformed cells is poor. Reductions in the frequency of stable integration events of 10 000-fold over the initial transfer event or transient expression have been reported following bombardment and *Agrobacterium*-mediated transformation (Russell *et al.* 1992a; Maximova *et al.* 1998). Techniques to manipulate bacterial virulence and optimize the physiological condition of the explant can increase the efficiency of transformation, but such manipulations often fail to enhance transformation in recalcitrant species – many of which are in fact able to induce *Agrobacterium* virulence. It was this line of thinking that led Gelvin (2000) to suggest that a greater understanding of the mechanisms of T-DNA integration would likely provide the next enhancements of transformation efficiency. Whilst we know much about the roles of the various VIR proteins, we lack a complete picture regarding their function – particularly during T-DNA integration – and their interactions with plant proteins involved in the stages of transformation.

4.8.2 Interaction of plant genes with the processes of transformation

We are only just beginning to identify plant genes/proteins involved in transformation, and often lack a detailed understanding of their role in the process. Studies conducted to date for *Agrobacterium*-mediated transformation have clearly shown a role for plant genes in all stages of the transformation process.

4.8.2.1 Cell attachment

Two *Arabidopsis* ecotypes with attenuated transformation were found to bind fewer *Agrobacterium* cells than do highly transformable ecotypes (Gelvin 2000). Gelvin (2000) and colleagues (Nam *et al.* 1999) screened a T-DNA insertion library of the *Arabidopsis* ecotype Ws for mutants that are resistant to *Agrobacterium* transformation (*rat*) on the basis of root explant transformation. Of the *rat* mutants identified, two showed deficiencies in *Agrobacterium*-binding, attributable to knockouts of genes whose proteins are either components of the cell wall or delivered to the apoplast.

4.8.2.2 Transfer and targeting

Targeting of the T-DNA to the plant nucleus must occur following transfer to the plant cell in order for stable integration to take place. Evidence indicates that the virulence proteins VIRD2 and VIRE2 interact with plant proteins during transfer and targeting of the T-strand to the nucleus. VIRD2 and VIRE2 contain nuclear localization signals (NLS) that function within plant cells (for a review, see Gelvin 2000).

VIRD2 contains two NLS sequences, although only one appears likely to be functional in T-DNA targeting as the second site is thought to be masked by the

linkage of the T-strand to the VIRD2 protein. VIRE2 also contains functional NLS sequences that are capable of targeting linked reporter proteins to the nucleus (Citovsky *et al.* 1992, 1994). Thus far, insufficient evidence has been obtained to determine the exact role of NLS sequences of either VIRD2 or VIRE2 in targeting the T-strand to the nucleus. Experiments in which NLS sequences have been deleted have shown attenuation of transformation (Rossi *et al.* 1993, 1996). Early results suggested that both VIRD2 and VIRE2 were required for nuclear targeting (Ziemienowicz *et al.* 1999), although VIRE2 does appear capable of targeting the T-strand to the nucleus in the absence of VIRD2 (Shurvinton *et al.* 1992; Mysore *et al.* 1998). The former studies utilized larger NLS mutations which may have affected other VIRD2 functions (Gelvin 2000).

Four plant proteins have been identified that play a role in nuclear targeting of the T-strand. An importin-α, which specifically binds NLS sequences in karyophilic proteins and assists in nuclear targeting, also binds to a NLS in VIRD2 (Ballas & Citovsky 1997). A plant serine/threonine protein phosphatase (PPC2) that negatively regulates nuclear localization of a β-glucuronidase (GUS)-VIRD2 fusion protein has been identified. Several lines of evidence indicate that expression of PPC2 appears to prevent nuclear localization through phosphorylation of a serine residue close to the NLS sequence in VIRD2 that is thought to be functional in targeting of the T-strand (see Gelvin 2000). An *Arabidopsis* cyclophilin has also been shown to interact with VIRD2 (Deng *et al.* 1998) and may act as a chaperonin to maintain conformation of the VID2 T-strand complex in a transport competent form.

Tzfira *et al.* (2001) have identified a plant protein that interacts specifically with VIRE2, called VIP1. This protein possesses two structural features, a basic domain and a leucine zipper, in common with basic-zipper (bZIP) proteins that are known to localize to the plant nucleus. VIP1 specifically assists in the nuclear import of VIRE2 and plants expressing anti-sense sequences to VIP1 show reduced susceptibility to *Agrobacterium*-mediated transformation. In the antisense *vip1* plants, transient expression of T-DNA genes is reduced as a consequence of impaired nuclear import of VIRE2, indicating that the block to transformation occurs at an early stage. Interestingly, nuclear import of VIRD2 was not affected in antisense plants. Cytoplasmic VIP1 is thought to interact specifically with VIRE2 and facilitate targeting to the nucleus via a karyopherin α-dependent pathway. It has been speculated that because the bZIP motif is found in many transcription factors that bind to the genomic DNA, VIP1 may also function to target the T-strand via VIRE2 to an integration site, though this hypothesis has yet to be confirmed. Nevertheless, over-expression of VIP1 increases both transient and stable transformation, most probably by increasing nuclear import of VIRE2 (Tzfira *et al.* 2002).

4.8.2.3 T-DNA integration

Usually, the rate of transient expression of T-DNA is much higher than that obtained following stable transformation, the inference being that much of the DNA entering the plant cell does not integrate into the genome. However, relatively little is known about the processes by which T-DNA integrates into the plant genome. Transient

expression occurs in the nucleus; single-stranded T-DNA is converted into a transcriptionally active form following the synthesis of a second strand. Consequently, much of the T-DNA entering the cell is converted to a double-stranded (ds) form. It is not clear whether this represents a dead-end molecule that is eventually lost or degraded, or if the ds T-DNA can be integrated into the genome. As Gelvin (2000) points out, it is not known whether T-DNA integration occurs by single-strand invasion of a locally denatured region, followed by second strand repair synthesis, or whether the T-strand is converted to an extrachromosomal double-stranded form prior to integration (DeNeve *et al.* 1997). Further studies are required to resolve the process of T-DNA integration. Clearly if ds-T-DNAs are a dead end for integration, then maximizing transient transformation events is unlikely to lead to increases in stable transformation. The results obtained by Cogan *et al.* (unpublished) following the genetic analysis of transformation are interesting in that lines containing QTL for efficient stable transformation also produce high rates of transient transformation events; this suggests that in these lines there is an efficient conversion of transient events to stable events. It remains to be shown whether these observations of high transient transformation leading to higher rates of stable integration are the result of ds-T-DNA integration or the consequence of other processes such as simply higher numbers of ss-T-DNAs reaching the nucleus and being available for either transient expression or stable integration.

There is evidence to indicate that the virulence proteins are involved in T-DNA integration. VIRE2 proteins protect the T-strand from nucleolytic degradation during transfer and targeting of the T-DNA and thus play an indirect role in integration. Mutations in the VIRD2 protein are known to affect both the precision of T-DNA integration at the 5' end of the T-strand and the efficiency of integration (reviewed in Gelvin 2000).

The role of plant proteins in the integration of T-DNA is only just beginning to be resolved. As previously stated, the mechanism of T-DNA integration is thought to be by illegitimate recombination (Matsumoto *et al.* 1990; Gheysen *et al.* 1991; Mayerhofer *et al.* 1991). This has led Gelvin (2000) and others to speculate that plants deficient in DNA repair and recombination might also be deficient for T-DNA integration. Plants that are sensitive to either ionizing radiation or ultraviolet light lack the ability to repair lesions induced in their DNA. Two mutants, *uvh1* and *rad5*, were initially found to be deficient for T-DNA integration following *Agrobacterium*-mediated transformation, although both transiently expressed transferred T-DNAs (Sonti *et al.* 1995). However, subsequent experiments using these mutants have led to a re-examination of the earlier findings. Using different *Agrobacterium* strains, T-DNA integration into the *uvh1* mutant line was established at efficiencies comparable with that of the wild-type line, whilst the *rad5* mutant was deficient for stable integration (Nam *et al.* 1998). The *rad5* mutant was also found to be attenuated for transient expression of genes carried on the T-DNA, leading to the conclusion that this mutant is probably blocked at a stage prior to T-DNA integration (Nam *et al.* 1998). Further radiation-sensitive ecotypes of *Arabidopsis* have been examined for their transformability. The ecotype UE-1 has been shown to be deficient in T-DNA

integration. In contrast to the recent findings with the *rad5* mutant, UE-1 shows very high levels of transient expression of T-DNA genes but very low levels of stable integration (Nam *et al.* 1998).

In addition to *Arabidopsis* ecotypes that are deficient for stable integration, several mutant (*rat*) lines have also been identified from the Ws T-DNA knock-out lines (Gelvin 2000). One of these mutant lines, *rat5*, has been characterized and found to carry a knock-out in the histone H2A gene (Mysore *et al.* 2000). It appears that disruption of this histone prevents insertion of T-DNA, possibly through modification of chromatin structure at the target site for insertion. According to Gelvin (2000), recent evidence shows that H2A interacts with VIRD2, suggesting that this might lead to the formation of an 'integration complex', though the exact mechanism has yet to be fully understood. Complementation of the *rat5* mutation with the wild-type H2A gene restores transformation efficiency and over-expression of the H2A gene has also been found to increase the efficiency of stable integration two- to three-fold.

The complexities of identifying plant genes that are involved in the processes of transformation are exacerbated by target tissue or explant factors within genotypes. In addition to the pronounced genotypic effects, specific cell types are also found to be more or less amenable to transformation than others. In the case of *rat5*, root explants are deficient for *Agrobacterium*-mediated transformation, but this line is susceptible to transformation via vacuum infiltration of the floral tissues (Mysore *et al.* 2000) where the ovule is the target (Ye *et al.* 1999). Similar differences in the susceptibility of different tissues have been found previously; in *Arabidopsis*, single T-DNA insertion events predominate in root explants and multiple events in leaf explants (Grevelding *et al.* 1993). The implication of these observations is that the plant genes/proteins which interact during the process of transformation may vary between different tissues, even within the same genotype. Consequently, the adventitious insertion of DNA into a plant genome may well exploit a variety of mechanisms which operate differentially in different cell types and at different stages of plant development.

4.9 Conclusions

The techniques of plant transformation have facilitated an expansion of our understanding of plant biology and provided the technology to modify and improve crop plants in novel ways. We are able address problems in plant breeding and plant biology that enable us to add value to our crop plants through the addition of new input and output traits as well as gaining further insights into plant growth and development.

A range of technologies has been developed for the modification of crop species, and in many cases several different approaches are available. The driving forces for the development of these transformation technologies are changing. Initial efforts were directed at overcoming recalcitrance or genotypic restrictions to the delivery and incorporation of DNA. An expansion in our knowledge of the molecular basis

of DNA transfer mediated by *Agrobacterium* has extended the range of plant species and genotypes that can be effectively transformed. This knowledge has also benefited some direct gene transfer techniques. Nevertheless, the genotypic differences revealed in the efficiency of gene transfer has suggested for some time that plant genes/proteins are also involved in determining the efficiency of the process. A clear genetic basis to transformation has been established, and classical breeding approaches have shown that responses to *Agrobacterium*-mediated transformation can be effectively manipulated. Often, the results of these studies have suggested that the basis of the differences observed is due to the action of recessive genes. We are at the early stages of identifying the plant genes/proteins that underlie these effects; nevertheless, we have recently gained some intriguing insights into the role of plant genes. Over-expression studies for VIP1 and H2A hold great promise for increasing the efficiency of transformation, provided that these results are replicated beyond the model plants (*Arabidopsis* and tobacco). The genetic factors regulating the efficiency of transformation can be studied and identified. In due course, further genes/proteins will be identified and this knowledge put to use in extending the efficiency of DNA transfer and integration. Emphasis is likely to focus on gaining greater predictability of gene transfer and subsequent gene expression.

References

Alonso, D.L. & Maroto, F.G. (2000) Plants as 'chemical factories' for the production of polyunsaturated fatty acids. *Biotechnology Advances*, **18**, 481–497.

Altmorbe, J., Kuhlmann, H. & Schroder, J. (1989) Differences in induction of Ti plasmid virulence genes virG and virD, and continued control of virD expression by four external factors. *Molecular Plant-Microbe Interactions*, **2**, 301–308.

Anderson, A. & Moore, L. (1979) Host specificity in the genus Agrobacterium. *Phytopathology*, **69**, 320–323.

Aoyama, T.& Chua, N.H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant Journal*, **11**, 605–612.

Asano, Y., Ito, Y., Fukami, M., Sugiura, K. & Fujiie, A. (1998) Herbicide-resistant transgenic creeping bentgrass plants obtained by electroporation using an altered buffer. *Plant Cell Reports*, **17**, 963–967.

Ballas, N. & Citovsky, V. (1997) Nuclear localization signal binding protein from Arabidopsis mediates nuclear import of Agrobacterium VirD2 protein. *Proceedings of the National Academy of Sciences, USA*, **94**, 10723–10728.

Barton., K.A., Binns, A.N., Matzke, A.J.M. & Chilton, M.D. (1983) Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T- DNA to R1 progeny. *Cell*, **32**, 1033–1043.

Bechtold, N., Ellis, J. & Pelletier, G. (1993) In-planta *Agrobacterium*-mediated gene-transfer by infiltration of adult *Arabidopsis thaliana* plants. *Comptes Rendus De L Academie Des Sciences Serie III, Sciences De La Vie-Life Sciences*, **316**, 1194–1199.

Bechtold, N., Jaudeau, B., Jolivet, S., *et al.* (2000) The maternal chromosome set is the target of the T-DNA in the in planta transformation of *Arabidopsis thaliana*. *Genetics*, **155**, 1875–1887.

Bendich, A.J. (1987) Why do chloroplasts and mitochondria contain so many copies of their genome. *BioEssays*, **6**, 279–282.

Beneddra, T., Picard, C., Petit, A. & Nesme, X. (1996) Correlation between susceptibility to crown gall and sensitivity to cytokinin in aspen cultivars. *Phytopathology*, **86**, 225–231.

Bent, A.F. (2000) *Arabidopsis* in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiology*, **124**, 1540–1547.

Bergmann, B.A. & Stomp, A.M. (1992) Effect of host plant genotype and growth rate on *Agrobacterium tumefaciens*-mediated gall formation in *Pinus radiata*. *Phytopathology*, **82**, 1457–1462.

Berthomieu, P., Beclin, C., Charlot, F., Dore, C. & Jouanin, L. (1994) Routine transformation of rapid cycling cabbage (*Brassica oleracea*) – molecular evidence for regeneration of chimeras. *Plant Science*, **96**, 223–235.

Berthomieu, P. & Jouanin, L. (1992) Transformation of rapid cycling cabbage (*Brassica oleracea* var. *capitata*) with *Agrobacterium rhizogenes*. *Plant Cell Reports*, **11**, 334–338.

Bidney, D., Scelonge, C., Martich, J., Burrus, M., Sims, L. & Huffman, G. (1992) Microprojectile bombardment of plant tissues increases transformation frequency by *Agrobacterium tumefaciens*. *Plant Molecular Biology*, **18**, 301–313.

Birch, R.G. (1997) Plant transformation: problems and strategies for practical application. *Annual Review of Plant Physiology and Plant Molecular Biology*, **48**, 297–326.

Bittencourt, P.A.L., Csanyi, A. & Jenes, B. (1995) Evaluation of different parameters and their influence on the PEG (polyethylene glycol) mediated gene transfer into rice (*Oryza sativa* L) protoplasts. *Cereal Research Communications*, **23**, 359–365.

Bouchez, D., Camilleri, C. & Caboche, M. (1993) A binary vector based on Basta resistance for in-planta transformation of *Arabidopsis thaliana*. *Comptes Rendus De L Academie Des Sciences Serie III,Sciences De La Vie-Life Sciences*, **316**, 1188–1193.

Boulter, M.E., Croy, E., Simpson, P., Shields, R., Croy, R.R.D. & Shirsat, A.H. (1990) Transformation of *Brassica napus* L (oilseed rape) using *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* – a comparison. *Plant Science*, **70**, 91–99.

Byrne, M.C., McDonnell, R.E., Wright, M.S. & Carnes, M.G. (1987) Strain and cultivar specificity in the *Agrobacterium*-soybean interaction. *Plant Cell Tissue Organ Culture*, **8**, 3–15.

Chandlee, J.M. (2001) Current molecular understanding of the genetically programmed process of leaf senescence. *Physiologia Planta*, **113**, 1–8.

Chang, S.S., Park, S.K., Kim, B.C., Kang, B.J., Kim, D.U. & Nam, H.G. (1994) Stable genetic transformation of Arabidopsis thaliana by Agrobacterium inoculation in planta. *Plant Journal*, **5**, 551–558.

Charest, P.J., Iyer, V.N. & Miki, B.L. (1989) Virulence of *Agrobacterium tumefaciens* strains with *Brassica napus* and *Brassica juncea*. *Plant Cell Reports*, **8**, 303–306.

Chateau, S., Sangwan, R.S. & Sangwan-Norreel, B.S. (2000) Competence of *Arabidopsis thaliana* genotypes and mutants for *Agrobacterium tumefaciens*-mediated gene transfer: role of phytohormones. *Journal of Experimental Botany*, **51**, 1961–1968.

Cheng, M., Fry, J.E., Pang, S.Z., *et al.* (1997) Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiology*, **115**, 971–980.

Chilton, M.D. (1993) Agrobacterium gene transfer – progress on a poor mans vector for maize. *Proceedings of the National Academy of Sciences, USA*, **90**, 3119–3120.

Chowrira, G.M., Akella, V. & Lurquin, P.F. (1995) Electroporation-mediated gene transfer into intact nodal meristems in planta: generating transgenic plants without in vitro tissue culture. *Molecular Biotechnology*, **3**, 17–23.

Christey, M.C. (1997) Transgenic crop plants using *Agrobacterium rhizogenes*-mediated transformation. In: *Hairy Roots: Culture and Applications* (ed. P.M. Doran), pp. 99–111. Harwood, Amsterdam.

Christey, M.C. (2001) Use of Ri-mediated transformation for production of transgenic plants. *In Vitro Cellular Developmental Biology-Plant*, **37**, 687–700.

Christie, P.J. & Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends in Microbiology*, **8**, 354–360.

Christou P (1995) Strategies for variety-independent genetic-transformation of important cereals, legumes and woody species utilizing particle bombardment. *Euphytica*, **85**, 13–27.

Christou, P. (1996) Transformation technology. *Trends in Plant Science*, **1**, 423–431.

Citovsky, V., Warnick, D. & Zambryski, P. (1994) Nuclear import of Agrobacterium VirD2 and VirE2 proteins in maize and tobacco. *Proceedings of the National Academy of Sciences, USA*, **12**, 3210–3214.

Citovsky, V., Zupan, J., Warnick, D. & Zambryski, P. (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. *Science*, **26**, 1802–1805.

Clough, S.J. & Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, **16**, 735–743.

Cogan, N., Harvey, E., Robinson, H., *et al*. (2001) The effects of anther culture and plant genetic background on *Agrobacterium rhizogenes*-mediated transformation of commercial cultivars and derived doubled-haploid *Brassica oleracea*. *Plant Cell Reports*, **20**, 755–762.

Cogan, N.O.I., Lynn, J.R., King, G.J., Kearsey, M.J., Newbury, H.J. & Puddephat, I.J. (2002) Identification of genetic factors controlling the efficiency of *Agrobacterium rhizogenes*-mediated transformation in *Brassica oleracea* by QTL analysis. *Theoretical and Applied Genetics*, **105**, 568–576.

Colby, S.M., Juncosa, A.M. & Meredith, C.P. (1991) Cellular differences in *Agrobacterium* susceptibility and regenerative capacity restrict the development of transgenic grapevines. *Journal of the American Society of Horticultural Science*, **116**, 356–361.

Crossway, A., Oakes, J.V., Irvine, J.M., Ward, B., Knauf, V.C. & Shewmaker, C.K. (1986) Integration of foreign DNA following microinjection of tobacco mesophyll protoplasts. *Molecular and General Genetics*, **202**, 179–185.

Croy, R.R.D. (1993) *Plant Molecular Biology Labfax*. Bios Scientific Publishers, Oxford.

Dale, E.C. & Ow, D.W. (1991) Gene transfer with subsequent removal of the selection gene from the host genome. *Proceedings of the National Academy of Sciences, USA*, **88**, 10558–10562.

Dale, P.J. & Irwin, J.A. (1995) The release of transgenic plants from containment, and the move towards their widespread use in agriculture. *Euphytica*, **85**, 425–431.

Daley, M., Knauf, V.C., Summerfelt, K.R. & Turner, J.C. (1998) Co-transformation with one *Agrobacterium tumefaciens* strain containing two binary plasmids as a method for producing marker-free transgenic plants. *Plant Cell Reports*, **17**, 489–496.

Damiani, F. & Arcioni, S. (1991) Transformation of *Medicago arborea* L. with an *Agrobacterium rhizogenes* binary vector carrying the hygromycin resistance gene. *Plant Cell Reports*, **10**, 300–303.

Daniell, H. & Varma, S. (1998) Chloroplast-transgenic plants: Panacea no! Gene containment yes! *Nature Biotechnology*, **16**, 602–602.

Datta, S.K., Datta, K., Soltanifar, N., Donn, G. & Potrykus, I. (1992) Herbicide-resistant *Indica* rice plants from IRRI breeding line IR72 after PEG-mediated transformation of protoplasts. *Plant Molecular Biology*, **20**, 619–629.

Davis, M.E., Miller, A.R. & Lineberger, R.D. (1991) Temporal competence for transformation of *Lycopersicon esculentum* (L-Mill) cotyledons by *Agrobacterium tumefaciens* – Relation to wound healing and soluble plant factors. *Journal of Experimental Botany*, **42**, 359–364.

Deblock, M. & Debrouwer, D. (1991) 2 T-DNAs co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. *Theoretical and Applied Genetics*, **82**, 257–263.

Deblock, M., Debrouwer, D. & Tenning, P. (1989) Transformation of *Brassica napus* and *Brassica oleracea* using *Agrobacterium tumefaciens* and the expression of the bar and neo genes in the transgenic plants. *Plant Physiology*, **91**, 694–701.

Deblock, M., Herreraestrella, L., Vanmontagu, M., Schell, J. & Zambryski, P. (1984) Expression of foreign genes in regenerated plants and in their progeny. *EMBO Journal*, **3**, 1681–1689.

Deblock, M., Schell, J. & Vanmontagu, M. (1985) Chloroplast transformation by *Agrobacterium tumefaciens*. *EMBO Journal*, **4**, 1367–1372.

DeFramond, A.J., Back, E.W., Chilton, W.S., Kayes, L. & Chilton, M.D. (1986) 2 unlinked T-DNAs can transform the same tobacco plant cell and segregate in the F1 generation. *Molecular and General Genetics*, **202**, 125–131.

Dekathen, A. & Jacobsen, H.J. (1995) Cell competence for Agrobacterium-mediated DNA transfer in *Pisum sativum* L. *Transgenic Research*, **4**, 184–191.

DeNeve, M., DeBuck, S., Jacobs, A., VanMontagu, M. & Depicker, A. (1997) T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs. *Plant Journal*, **11**, 15–29.

Deng, W., Chen, L., Metcalfe, T., *et al*. (1998) *Agrobacterium* VirD2 protein interacts with plant host cyclophilins. *Proceedings of the National Academy of Sciences, USA*, **95**, 7040–7045.

Depicker, A., Herman, L., Jacobs, A., Schell, J. & Vanmontagu, M. (1985) Frequencies of simultaneous transformation with different T- DNAs and their relevance to the *Agrobacterium* plant cell interaction. *Molecular and General Genetics*, **201**, 477–484.

Desfeux, C., Clough, S.J. & Bent, A.F. (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiology*, **123**, 895–904.

Dhir, S.K., Dhir, S., Sturtevant, A.P. & Widholm, J.M. (1991) Regeneration of transformed shoots from electroporated soybean (*Glycine max* (L) Merr) protoplasts. *Plant Cell Reports*, **10**, 97–101.

Dillen, W., De Clercq, J., Kapila, J., Zambre, M., Van Montagu, M. & Angenon, G. (1997) The effect of temperature on *Agrobacterium tumefaciens*-mediated gene transfer to plants. *Plant Journal*, **12**, 1459–1463.

Dommisse, E.M., Leung, D.W.M., Shaw, M.L. & Conner, A.J. (1990) Onion is a monocotyledonous host for *Agrobacterium*. *Plant Science*, **69**, 249–257.

Draper, J., Scott, R., Armitage, P. & Walden, R. (1988) *Plant Genetic Transformation and Gene Expression: A Laboratory Manual*. Blackwell Scientific Publications, Oxford.

Ebinuma, H. & Komamine, A. (2001) MAT (Multi-Auto-Transformation) Vector System. The oncogenes of Agrobacterium as positive markers for regeneration and selection of marker-free transgenic plants. *In Vitro Cellular Development Biology-Plant*, **37**, 103–113.

Ebinuma, H., Sugita, K., Matsunaga, E., Endo, S., Yamada, K. & Komamine, A. (2001) Systems for the removal of a selection marker and their combination with a positive marker. *Plant Cell Reports*, **20**, 383–392.

Enriquez-Obregon, G.A., Prieto-Samsonov, D.L., de la Riva, G.A., Perez, M., Selman-Housein, G. & Vazquez-Padron, R.I. (1999) *Agrobacterium*-mediated Japonica rice transformation: a procedure assisted by an antinecrotic treatment. *Plant Cell Tissue Organ Culture*, **59**, 159–168.

Escudero, J. & Hohn, B. (1997) Transfer and integration of T-DNA without cell injury in the host plant. *Plant Cell*, **9**, 2135–2142.

Escudero, J., Neuhaus, G. & Hohn, B. (1995) Intracellular *Agrobacterium* can transfer DNA to the cell-nucleus of the host-plant. *Proceedings of the National Academy of Sciences, USA*, **92**, 230–234.

Escudero, J., Neuhaus, G., Schlappi, M. & Hohn, B. (1996) T-DNA transfer in meristematic cells of maize provided with intracellular *Agrobacterium*. *Plant Journal*, **10**, 355–360.

Evans, I.J. & Greenland, A.J. (1998) Transgenic approaches to disease protection: applications of antifungal proteins. *Pesticide Science*, **54**, 353–359.

Feldmann, K.A. & Marks, M.D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. *Molecular and General Genetics*, **208**, 1–9.

Fischer, R. & Emans, N. (2000) Molecular farming of pharmaceutical proteins. *Transgenic Research*, **9**, 279–299.

Fischer, R. & Hain, R. (1995) Tobacco protoplast transformation and use for functional analysis of newly isolated genes and gene constructs. In: *Methods in Cell Biology*, pp. 401–410. Academic Press Inc., San Diego, CA.

Fischer, R., Schillberg, S. & Emans, N. (2001) Molecular farming of medicines: a field of growing promise. *Outlook in Agriculture*, **30**, 31–36.

Flavell, R.B., Dart, E., Fuchs, R.L. & Fraley, R.T. (1992) Selectable marker genes – Safe for plants. *Bio-Technology*, **10**, 141–144.

Fraley, R.T., Rogers, S.G., Horsch, R.B., *et al*. (1983) Expression of bacterial genes in plant cells. *Proceedings of the National Academy of Sciences, USA*, **80**, 4803–4807.

Fransz, P.F., Stam, M., Montijn, B., *et al*. (1996) Detection of single-copy genes and chromosome rearrangements in *Petunia hybrida* by fluorescence in situ hybridization. *Plant Journal*, **9**, 767–774.

Fromm, M., Taylor, L.P. & Walbot, V. (1985) Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proceedings of the National Academy of Sciences, USA*, **82**, 5824–5828.

Fuchs, R.L., Ream, J.E., Hammond, B.G., Naylor, M.W., Leimgruber, R.M. & Berberich, S.A. (1993) Safety assessment of the neomycin phosphotransferase-II (NPTII) protein. *Bio-Technology*, **11**, 1543–1547.

Fullner, K.J., Lara, J.C. & Nester, E.W. (1996) Pilus assembly by *Agrobacterium* T-DNA transfer genes. *Science*, **273**, 1107–1109.

Fullner, K.J. & Nester, E.W. (1996) Temperature affects the T-DNA transfer machinery of *Agrobacterium tumefaciens. Journal of Bacteriology*, **178**, 1498–1504.

Gelvin, S.B. (1990) Crown gall disease and hairy root disease – a sledgehammer and a tackhammer. *Plant Physiology*, **92**, 281–285.

Gelvin, S.B. (1998) The introduction and expression of transgenes in plants. *Current Opinion in Biotechnology*, **9**, 227–232.

Gelvin, S.B. (2000) Agrobacterium and plant genes involved in T-DNA transfer and integration. *Annual Review Plant Physiology and Plant Molecular Biology*, **51**, 223–256.

Gelvin, S.B. & Schilperoort, R.A. (1988) *Plant Molecular Biology Manual*. Kluwer, Dordrecht, Netherlands.

Gheysen, G., Villarroel, R. & Vanmontagu, M. (1991) Illegitimate recombination in plants – a model for T-DNA integration. *Genes and Development*, **5**, 287–297.

Gilbertson, L., Addae, P., Ekena, J., *et al*. (2002) Cre/Lox mediated marker gene excision in transgenic crop plants. In: Plant biotechnology 2002 and Beyond. (ed. I.K. Vasil), pp. 225–228. Proceedings of the 10th IAPTC&B Congress. Kluwer Academic Publishers, Orlando, Florida.

Gleave, A.P., Mitra, D.S., Mudge, S.R. & Morris, B.A.M. (1999) Selectable marker-free transgenic plants without sexual crossing: transient expression of cre recombinase and use of a conditional lethal dominant gene. *Plant Molecular Biology*, **40**, 223–235.

Godwin, I., Todd, G., Fordlloyd, B.V. & Newbury, H.J. (1991) The effects of acetosyringone and pH on *Agrobacterium*-mediated transformation vary according to plant-species. *Plant Cell Reports*, **9**, 671–675.

Goldsbrough, A.P., Lastrella, C.N. & Yoder, J.I. (1993) Transposition mediated repositioning and subsequent elimination of marker genes from transgenic tomato. *Bio-Technology*, **11**, 1286–1292.

Grayburn, W.S. & Vick, B.A. (1995) Transformation of sunflower (*Helianthus annuus* L) following wounding with glass-beads. *Plant Cell Reports*, **14**, 285–289.

Grevelding, C., Fantes, V., Kemper, E., Schell, J. & Masterson, R. (1993) Single-copy T-DNA insertions in *Arabidopsis* are the predominant form of integration in root-derived transgenics, whereas multiple insertions are found in leaf discs. *Plant Molecular Biology*, **23**, 847–860.

Grierson D (1991) *Plant Genetic Engineering. Plant Biotechnology, 1.* Blackie, Glasgow.

Guo, G.Q., Maiwald, F., Lorenzen, P. & Steinbiss, H.H. (1998) Factors influencing T-DNA transfer into wheat and barley cells by *Agrobacterium tumefaciens. Cereal Research Communications*, **26**, 15–22.

Haldrup, A., Petersen, S.G. & Okkels, F.T. (1998) The xylose isomerase gene from *Thermoanaerobacterium thermosulfurogenes* allows effective selection of transgenic plant cells using D-xylose as the selection agent. *Plant Molecular Biology*, **37**, 287–296.

Hall, R.D., RiksenBruinsma, T., Weyens, G., *et al*. (1997) Sugar beet guard cell protoplasts demonstrate a remarkable capacity for cell division enabling applications in stomatal physiology and molecular breeding. *Journal of Experimental Botany*, **48**, 255–263.

Hall, R.D., RiksenBruinsma, T., Weyens, G.J., *et al*. (1996) A high efficiency technique for the generation of transgenic sugar beets from stomatal guard cells. *Nature Biotechnology*, **14**, 1133–1138.

Hamill, J.D. & Lidgett, A.J. (1997) Hairy root cultures: opportunities and key protocols for studies in metabolic engineering. In: *Hairy Roots: Culture and Applications* (ed. P.M. Doran), pp. 1–29.Harwood Academic Publishers, Amsterdam.

Hamilton, C.M., Frary, A., Lewis, C. & Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proceedings of the National Academy of Sciences, USA*, **93**, 9975–9979.

Han, K.H., Keathley, D.E., Davis, J.M. & Gordon, M.P. (1993) Regeneration of a transgenic woody legume (*Robinia pseudoacacia* L, Black Locust) and morphological alterations induced by *Agrobacterium rhizogenes*-mediated transformation. *Plant Science*, **88**, 149–157.

Hansen, G. & Chilton, M.D. (1996) ''Agrolistic'' transformation of plant cells: integration of T- strands generated in planta. *Proceedings of the National Academy of Sciences, USA*, **93**, 14978–14983.

Hansen, G. & Wright, M.S. (1999) Recent advances in the transformation of plants. *Trends in Plant Science*, **4**, 226–231.

Hatamoto, H., Boulter, M.E., Shirsat, A.H., Croy, E.J. & Ellis, J.R. (1990) Recovery of morphologically normal transgenic tobacco from hairy roots co-transformed with *Agrobacterium rhizogenes* and a binary vector plasmid. *Plant Cell Reports*, **9**, 88–92.

Hellens, R., Mullineaux, P. & Klee, H. (2000a) A guide to *Agrobacterium* binary Ti vectors. *Trends in Plant Science*, **5**, 446–451.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. & Mullineaux, P.M. (2000b) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Molecular Biology*, **42**, 819–832.

Henzi, M.X., Christey, M.C. & McNeil, D.L. (2000) Factors that influence *Agrobacterium rhizogenes*-mediated transformation of broccoli (*Brassica oleracea* L. var. *italica*). *Plant Cell Reports*, **19**, 994–999.

Herbers, K. & Sonnewald, U. (1996) Manipulating metabolic partitioning in transgenic plants. *Trends in Biotechnology*, **14**, 198–205.

Herbers, K. & Sonnewald, U. (1998) Molecular determinants of sink strength. *Current Opinion in Plant Biology*, **1**, 207–216.

Herrera-Estrella, L., Depicker, A., Vanmontagu, M. & Schell, J. (1983) Expression of chimaeric genes transferred into plant cells using a Ti-plasmid derived vector. *Nature*, **303**, 209–213.

Hess, D., Dressler, K. & Nimmrichter, R. (1990) Transformation experiments by pipetting *Agrobacterium* into the spikelets of wheat (*Triticum aestivum* L). *Plant Science*, **72**, 233–244.

Hiei, Y., Komari, T. & Kubo, T. (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Molecular Biology*, **35**, 205–218.

Holm, P.B., Olsen, O., Schnorf, M., Brinch-Pedersen, H. & Knudsen, S. (2000) Transformation of barley by microinjection into isolated zygote protoplasts. *Transgenic Research*, **9**, 21–32.

Hood, E.E., Fraley, R.T. & Chilton, M.D. (1987) Virulence of *Agrobacterium tumefaciens* strain A281 on legumes. *Plant Physiology*, **83**, 529–534.

Hoopen, R.T., Montijn, B.M., Veuskens, J.T.M., Oud, O.J.L. & Nanninga, N. (1999) The spatial localization of T-DNA insertions in petunia interphase nuclei: consequences for chromosome organization and transgene insertion sites. *Chromosome Research*, **7**, 611–623.

Hoopen, R.T., Robbins, T.P., Fransz, P.F., *et al*. (1996) Localization of T-DNA insertions in petunia by fluorescence in situ hybridization: physical evidence for suppression of recombination. *Plant Cell*, **8**, 823–830.

Hooykaas, P.J.J. & Beijersbergen, A.G.M. (1994) The virulence system of *Agrobacterium tumefaciens. Annual Review of Phytopathology*, **32**, 157–179.

Hooykaas, P.J.J. & Schilperoort, R.A. (1992) *Agrobacterium* and plant genetic engineering. *Plant Molecular Biology*, **19**, 15–38.

Hooykaas, P.J.J., Hofker, M., Dendulkras, H. & Schilperoort, R.A. (1984) A comparison of virulence determinants in an octopine Ti- plasmid, a nopaline Ti-plasmid, and an Ri-plasmid by complementation analysis of *Agrobacterium tumefaciens* mutants. *Plasmid*, **11**, 195–205.

Horsch, R.B., Fraley, R.T., Rogers, S.G., Sanders, P.R., Lloyd, A. & Hoffmann, N. (1984) Inheritance of functional foreign genes in plants. *Science*, **223**, 496–498.

Hu, C.-Y. & Wang, L. (1999) In planta transformation technologies developed in China: procedure, confirmation and field performance. *In Vitro Cellular Development Biol-Plant*, **35**, 417–420.

Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T. & Kumashiro, T. (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens. Nature Biotechnology*, **14**, 745–750.

James, C. (1997) Global status of transgenic crops in 1997. ISAAA, Ithaca, NY pp. 31.

James, C. (2001) ISAAA Briefs, No. 24.

Jefferson, R.A., Kavanagh, T.A. & Bevan, M.W. (1987) GUS fusions: beta-glucuronidase (GUS) as a sensitive and versatile gene fusion marker in plants. *EMBO Journal*, **6**, 3901–3907.

Jenes, B., Moore, H., Cao, J., Zhang, W. & Wu, R. (1993) Techniques for gene transfer. In: *Transgenic Plants, Volume 1* (ed. S. Kung and R. Wu), pp. 125–146. Academic Press, San Diego.

Joersbo, M. (2001) Advances in the selection of transgenic plants using non-antibiotic marker genes. *Physiologia Planta*, **111**, 269–272.

Joersbo, M. & Okkels, F.T. (1996) A novel principle for selection of transgenic plant cells: positive selection. *Plant Cell Reports*, **16**, 219–221.

Jones-Villeneuve, E., Huang, B., Prudhomme, I., *et al*. (1995) Assessment of microinjection for introducing DNA into uninuclear microspores of rapeseed. *Plant Cell Tissue Organ Culture*, **40**, 97–100.

Jordan, M.C. & McHughen, A. (1988) Transformed callus does not necessarily regenerate transformed shoots. *Plant Cell Reports*, **7**, 285–287.

Kado, C.I. (1991) Molecular mechanisms of crown gall tumorigenesis. *Critical Reviews in Plant Science*, **10**, 1–32.

Kado, C.I. (1994) Promiscuous DNA transfer system of *Agrobacterium tumefaciens*: role of the Virb operon in sex pilus assembly and synthesis. *Molecular Microbiology*, **12**, 17–22.

Katavic, V., Haughn, G.W., Reed, D., Martin, M.& Kunst, L. (1994) In planta transformation of *Arabidopsis thaliana*. *Molecular and General Genetics*, **245**, 363–370.

Katavicacute, V. & Jelaska, S. (1991) The influence of plant growth regulators on callus induction in pumpkin (*Cucurbita pepo* L.) hairy roots. *International Journal of Developmental Biology*, **35**, 265–268.

Ke, J., Khan, R., Johnson, T., Somers, D.A. & Das, A. (2001) High-efficiency gene transfer to recalcitrant plants by *Agrobacterium tumefaciens. Plant Cell Reports*, **20**, 150–156.

Khanna, H.K. & Raina, S.K. (1999) Agrobacterium-mediated transformation of indica rice cultivars using binary and superbinary vectors. *Australian Journal of Plant Physiology*, **26**, 311–324.

Khrustaleva, L.I. & Kik, C. (2001) Localization of single-copy T-DNA insertion in transgenic shallots (*Allium cepa*) by using ultra-sensitive FISH with tyramide signal amplification. *Plant Journal*, **25**, 699–707.

Kikkert, J.R. (1993) The biolistic PDS-1000 He device. *Plant Cell Tissue Organ Culture*, **33**, 221–226.

Kikkert, J.R., Humiston, G.A., Roy, M.K. & Sanford, J.C. (1999) Biological projectiles (phage, yeast, bacteria) for genetic transformation of plants. *In Vitro Cellular and Developmental Biology-Plant*, **35**, 43–50.

Kohli, A., Leech, M., Vain, P., Laurie, D.A. & Christou, P. (1998) Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. *Proceedings of the National Academy of Sciences, USA*, **95**, 7203–7208.

Komari, T., Hiei, Y., Saito, Y., Murai, N. & Kumashiro, T. (1996) Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by *Agrobacterium tumefaciens* and segregation of transformants free from selection markers. *Plant Journal*, **10**, 165–174.

Koop, H.U., Steinmuller, K., Wagner, H., Rossler, C., Eibl, C. & Sacher, L. (1996) Integration of foreign sequences into the tobacco plastome via polyethylene glycol-mediated protoplast transformation. *Planta*, **199**, 193–201.

Kost, B., Galli, A., Potrykus, I. & Neuhaus, G. (1995) High efficiency transient and stable transformation by optimized DNA microinjection into *Nicotiana tabacum* protoplasts. *Journal of Experimental Botany*, **46**, 1157–1167.

Kramer, K.J. & Muthukrishnan, S. (1997) Insect chitinases: molecular biology and potential use as biopesticides. *Insect Biochemistry and Molecular Biology*, **27**, 887–900.

Kumpatla, S.P., Chandrasekharan, M.B., Iyer, L.M., Li, G.F. & Hall, T.C. (1998) Genome intruder scanning and modulation systems and transgene silencing. *Trends in Plant Science*, **3**, 97–104.

Kung, S. & Wu, R. (1993) *Transgenic Plants*. Academic Press, London.

Kunkel, T., Niu, Q.W., Chan, Y.S. & Chua, N.H. (1999) Inducible isopentenyl transferase as a high-efficiency marker for plant transformation. *Nature Biotechnology*, **17**, 916–919.

Kunze, I., Ebneth, M., Heim, U., Geiger, M., Sonnewald, U. & Herbers, K. (2001) 2-Deoxyglucose resistance: a novel selection marker for plant transformation. *Molecular Breeding*, **7**, 221–227.

Leduc, N., MatthysRochon, E., Rougier, M., *et al*. (1996) Isolated maize zygotes mimic in vivo embryonic development and express microinjected genes when cultured in vitro. *Developmental Biology*, **177**, 190–203.

Lee, L.Y., Humara, J.M. & Gelvin, S.B. (2001) Novel constructions to enable the integration of genes into the *Agrobacterium tumefaciens* C58 chromosome. *Molecular Plant-Microbe Interactions*, **14**, 577–579.

Lelievre, J.M., Latche, A., Jones, B., Bouzayen, M. & Pech, J.C. (1997) Ethylene and fruit ripening. *Physiologia Planta*, **101**, 727–739.

Li, X.Q., Liu, C.N., Ritchie, S.W., Peng, J.Y., Gelvin, S.B. & Hodges, T.K. (1992) Factors influencing *Agrobacterium*-mediated transient expression of GUSA in rice. *Plant Molecular Biology*, **20**, 1037–1048.

Lowe, K., Bowen, B., Hoerster, G., *et al*. (1995) Germline transformation of maize following manipulation of chimeric shoot meristems. *Bio-Technology*, **13**, 677–682.

Lydiate, D., Dale, P., Lagercrantz, U., Parkin, I. & Howell, P. (1995) Selecting the optimum genetic background for transgenic varieties, with examples from *Brassica*. *Euphytica*, **85**, 351–358.

Maliga, P. (2001) Plastid engineering bears fruit – The tomato has been engineered to express in its plastids high levels of a recombinant protein. *Nature Biotechnology*, **19**, 826–827.

Manners, J.M. & Way, H. (1989) Efficient transformation with regeneration of the tropical pasture legume *Stylosanthes humilis* using *Agrobacterium rhizogenes* and a Ti plasmid-binary vector system. *Plant Cell Reports*, **8**, 341–345.

Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y. & Machida, Y. (1990) Integration of *Agrobacterium* T-DNA into a tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. *Molecular and General Genetics*, **224**, 309–316.

Matzk, A., Mantell, S. & Schiemann, J. (1996) Localization of persisting Agrobacteria in transgenic tobacco plants. *Molecular Plant-Microbe Interactions*, **9**, 373–381.

Matzke, A.J.M. & Matzke, M.A. (1995) Trans-inactivation of homologous sequences in *Nicotiana tabacum*. In: *Gene Silencing in Higher Plants and Related Phenomena in other Eukaryotes, Volume 197*, pp. 1–14. Springer-Verlag, Berlin.

Mauro, A.O., Nobrega, J.C., Baraldi, G.L. & Collins, G.B. (1995a) Susceptibility of some Brazilian soybean genotypes to 3 strains of *Agrobacterium tumefaciens. Revista Brasileira De Genetica*, **18**, 417–420.

Mauro, A.O., Pfeiffer, T.W. & Collins, G.B. (1995b) Inheritance of soybean susceptibility to *Agrobacterium tumefaciens* and its relationship to transformation. *Crop Science*, **35**, 1152–1156.

Maximova, S.N., Dandekar, A.M. & Guiltinan, M.J. (1998) Investigation of *Agrobacterium*-mediated transformation of apple using green fluorescent protein: high transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting. *Plant Molecular Biology*, **37**, 549–559.

May, G.D., Afza, R., Mason, H.S., Wiecko, A., Novak, F.J. & Arntzen, C.J. (1995) Generation of transgenic Banana (*Musa acuminata*) plants via *Agrobacterium*-mediated transformation. *Bio-Technology*, **13**, 486–492.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., *et al*. (1991) T-DNA integration: a mode of illegitimate recombination in plants. *EMBO Journal*, **10**, 697–704.

McBride, K.E., Schaaf, D.J., Daley, M. & Stalker, D.M. (1994) Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proceedings of the National Academy of Sciences, USA*, **91**, 7301–7305.

McBride, K.E., Svab, Z., Schaaf, D.J., Hogan, P.S., Stalker, D.M. & Maliga, P. (1995) Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *Bio-Technology*, **13**, 362–365.

McCormac, A.C., Fowler, M.R., Chen, D.F. & Elliott, M.C. (2001) Efficient co-transformation of *Nicotiana tabacum* by two independent T-DNAs, the effect of T-DNA size and implications for genetic separation. *Transgenic Research*, **10**, 143–155.

McKnight, T.D., Lillis, M.T. & Simpson, R.B. (1987) Segregation of genes transferred to one plant cell from two separate *Agrobacterium* strains. *Plant Molecular Biology*, **8**, 439–445.

Medford, J.I., Horgan, R., Elsawi, Z. & Klee, H.J. (1989) Alterations of endogenous cytokinins in transgenic plants using a chimeric isopentenyl transferase gene. *Plant Cell*, **1**, 403–413.

Mol, J.N.M., Holton, T.A. & Koes, R.E. (1995) Floriculture – genetic engineering of commercial traits. *Trends in Biotechnology*, **13**, 350–355.

Mukhopadhyay, A., Topfer, R., Pradhan, A.K., *et al*. (1991) Efficient regeneration of *Brassica oleracea* hypocotyl protoplasts and high-frequency genetic transformation by direct DNA uptake. *Plant Cell Reports*, **10**, 375–379.

Mysore, K.S., Bassuner, B., Deng, X.B., *et al*. (1998) Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. *Molecular Plant-Microbe Interactions*, **11**, 668–683.

Mysore, K.S., Nam, J. & Gelvin, S.B. (2000) An *Arabidopsis* histone H2A mutant is deficient in *Agrobacterium* T-DNA integration. *Proceedings of the National Academy of Sciences, USA*, **97**, 948–953.

Nadolska-Orczyk, A. & Orczyk, W. (2000) Study of the factors influencing *Agrobacterium*-mediated transformation of pea (*Pisum sativum* L.). *Molecular Breeding*, **6**, 185–194.

Nam, J., Matthysse, A.G. & Gelvin, S.B. (1997) Differences in susceptibility of *Arabidopsis* ecotypes to crown gall disease may result from a deficiency in T-DNA integration. *Plant Cell*, **9**, 317–333.

Nam, J., Mysore, K.S. & Gelvin, S.B. (1998) *Agrobacterium tumefaciens* transformation of the radiation hypersensitive *Arabidopsis thaliana* mutants uvh1 and rad5. *Molecular Plant-Microbe Interactions*, **11**, 1136–1141.

Nam, J., Mysore, K.S., Zheng, C., Knue, M.K., Matthysse, A.G. & Gelvin, S.B. (1999) Identification of T-DNA tagged *Arabidopsis* mutants that are resistant to transformation by *Agrobacterium*. *Molecular and General Genetics*, **261**, 429–438.

Narasimhulu, S.B., Deng, X., Sarria, R. & Gelvin, S.B. (1996) Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. *Plant Cell*, **8**, 873–886.

Newbigin, E., Smyth, D.R. & Clarke, A.E. (1995) Understanding and controlling plant development. *Trends in Biotechnology*, **13**, 338–343.

Ohlsson, M. & Eriksson, T. (1988) Transformation of *Brassica campestris* protoplasts with *Agrobacterium tumefaciens*. *Hereditas*, **108**, 173–177.

Paul, M.J. & Foyer, C.H. (2001) Sink regulation of photosynthesis. *Journal of Experimental Botany*, **52**, 1383–1400.

Pawlowski, W.P. & Somers, D.A. (1996) Transgene inheritance in plants genetically engineered by microprojectile bombardment. *Molecular Biotechnology*, **6**, 17–30.

Pedersen, C., Zimny, J., Becker, D., JahneGartner, A. & Lorz, H. (1997) Localization of introduced genes on the chromosomes of transgenic barley, wheat and triticale by fluorescence in situ hybridization. *Theoretical and Applied Genetics*, **94**, 749–757.

Perl, A., Lotan, O., AbuAbied, M. & Holland, D. (1996) Establishment of an *Agrobacterium*-mediated transformation system for grape (*Vitis vinifera* L): the role of antioxidants during grape-*Agrobacterium* interactions. *Nature Biotechnology*, **14**, 624–628.

Pink, D. & Puddephat, I. (1999) Deployment of disease resistance genes by plant transformation – a 'mix and match' approach. *Trends in Plant Science*, **4**, 71–75.

Porter, J.R. (1991) Host range and implications of plant infection by *Agrobacterium rhizogenes*. *Critical Reviews in Plant Science*, **10**, 387–421.

Potrykus, I. (1991) Gene-transfer to plants – assessment of published approaches and results. *Annual Review of Plant Physiology and Plant Molecular Biology*, **42**, 205–225.

Potrykus, I. & Spangenberg, G. (1995) *Gene Transfer to Plants.* Springer Laboratory Manual, Berlin.

Puddephat, I.J., Robinson, H.T., Fenning, T.M., Barbara, D.J., Morton, A. & Pink, D.A.C. (2001) Recovery of phenotypically normal transgenic plants of *Brassica oleracea* upon *Agrobacterium rhizogenes*-mediated co-transformation and selection of transformed hairy roots by GUS assay. *Molecular Breeding*, **7**, 229–242.

Qing, C.M., Fan, L., Lei, Y., *et al.* (2000) Transformation of Pakchoi (*Brassica rapa* L. ssp. *chinensis*) by *Agrobacterium* infiltration. *Molecular Breeding*, **6**, 67–72.

Rao, K.V., Rathore, K.S. & Hodges, T.K. (1995) Physical, chemical and physiological parameters for electroporation-mediated gene delivery into rice protoplasts. *Transgenic Research*, **4**, 361–368.

Rasmussen, J.O. & Rasmussen, O.S. (1993) PEG mediated DNA uptake and transient GUS expression in carrot, rapeseed and soybean protoplasts. *Plant Science*, **89**, 199–207.

Reed, J., Privalle, L., Powell, M.L., *et al.* (2001) Phosphomannose isomerase: an efficient selectable marker for plant transformation. *In Vitro Cellular and Developmental Biology-Plant*, **37**, 127–132.

Robaglia, C., Vilaine, F., Pautot, V., *et al.* (1987) Expression vectors based on the *Agrobacterium rhizogenes* Ri plasmid transformation system. *Biochimie*, **69**, 231–237.

Robbs, S.L., Pueppke, S.G. & Hawes, M.C. (1987) Genotypic variation in susceptibility of *Pisum sativum* to infection by *Agrobacterium tumefaciens*. *Phytopathology*, **77**, 1241–1242.

Robinette, D. & Matthysse, A.G. (1990) Inhibition by *Agrobacterium tumefaciens* and *Pseudomonas savastanoi* of development of the hypersensitive response elicited by *Pseudomonas syringae* pv. *phaseolicola*. *Journal of Bacteriology*, **172**, 5742–5749.

Rogers, S.G., Horsch, R.B. & Fraley, R.T. (1986) Improved vectors for plant transformation: expression cassette vectors and new selectable markers. *Methods in Enzymology*, **118**, 627–641.

Rossi, L., Hohn, B. & Tinl, B. (1993) The VirD2 protein of *Agrobacterium tumefaciens* carries nuclear localization signals important for transfer of T-DNA to plants. *Molecular and General Genetics*, **239**, 345–353.

Rossi, L., Hohn, B. & Tinl, B. (1996) Integration of complete transferred DNA units is dependent on the activity of virulence E2 protein of *Agrobacterium tumefaciens. Proceedings of the National Academy of Sciences, USA*, **93**, 126–130.

Royal Society (1998) *Genetically Modified Plants for Food Use*. Royal Society, London.

Ruf, S., Hermann, M., Berger, I.J., Carrer, H. & Bock, R. (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nature Biotechnology*, **19**, 870–875.

Russell, J.A., Roy, M.K. & Sanford, J.C. (1992a) Physical trauma and tungsten toxicity reduce the efficiency of biolistic transformation. *Plant Physiology*, **98**, 1050–1056.

Russell, S.H., Hoopes, J.L. & Odell, J.T. (1992b) Directed excision of a transgene from the plant genome. *Molecular and General Genetics*, **234**, 49–59.

Salas, M.G., Park, S.H., Srivatanakul, M. & Smith, R.H. (2001) Temperature influence on stable T-DNA integration in plant cells. *Plant Cell Reports*, **20**, 701–705.

Salmeron, J.M. & Vernooij, B. (1998) Transgenic approaches to microbial disease resistance in crop plants. *Current Opinion in Plant Biology*, **1**, 347–352.

Salvo-Garrido, H., Travella, S., Schwarzacher, T., Harwood, W.A. & Snape, J.W. (2001) An efficient method for the physical mapping of transgenes in barley using in situ hybridization. *Genome*, **44**, 104–110.

Sanford, J.C., Smith, F.D. & Russell, J.A. (1993) Optimizing the biolistic process for different biological applications. *Methods in Enzymology*, **217**, 483–509.

Sangwan, R.S., Bourgeois, Y., Brown, S., Vasseur, G. & Sangwannorreel, B. (1992) Characterization of competent cells and early events of *Agrobacterium*-mediated genetic transformation in *Arabidopsis thaliana*. *Planta*, **188**, 439–456.

Santarem, E.R., Trick, H.N., Essig, J.S. & Finer, J.J. (1998) Sonication assisted *Agrobacterium*-mediated transformation of soybean immature cotyledons: optimization of transient expression. *Plant Cell Reports*, **17**, 752–759.

Sautter, C., Leduc, N., Bilang, R., *et al*. (1995) Shoot apical meristems as a target for gene-transfer by microballistics. *Euphytica*, **85**, 45–51.

Schmulling, T. & Schell, J. (1993) Transgenic tobacco plants regenerated from leaf-disks can be periclinal chimeras. *Plant Molecular Biology*, **21**, 705–708.

Schmulling, T., Schell, J. & Spena, A. (1988) Single genes from *Agrobacterium rhizogenes* influence plant development. *EMBO Journal,* **7**, 2621–2629.

Scholten, S. & Kranz, E. (2001) In vitro fertilization and expression of transgenes in gametes and zygotes. *Sexual Plant Reproduction*, **14**, 35–40.

Shah, D.M. (1997) Genetic engineering for fungal and bacterial diseases. *Current Opinion in Biotechnology*, **8**, 208–214.

Sheng, J.S. & Citovsky, V. (1996) Agrobacterium plant cell DNA transport: have virulence proteins, will travel. *Plant Cell*, **8**, 1699–1710.

Shurvinton, C.E., Hodges, L. & Ream, W. (1992) A nuclear localization signal and the C-terminal omega sequence in the *Agrobacterium tumefaciens* VirD2 endonuclease are important for tumor formation. *Proceedings of the National Academy of Sciences, USA*, **89**, 11837–11841.

Sidorov, V.A., Kasten, D., Pang, S.Z., Hajdukiewicz, P.T.J., Staub, J.M. & Nehra, N.S. (1999) Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant Journal*, **19**, 209–216.

Smigocki, A.C. & Owens, L.D. (1988) Cytokinin gene fused with a strong promoter enhances shoot organogenesis and zeatin levels in transformed plant cells. *Proceedings of the National Academy of Sciences, USA*, **85**, 5131–5135.

Smith, N., Kilpatrick, J.B. & Whitelam, G.C. (2001) Superfluous transgene integration in plants. *Critical Reviews in Plant Science*, **20**, 215–249.

Songstad, D.D., Somers, D.A. & Griesbach, R.J. (1995) Advances in alternative DNA delivery techniques. *Plant Cell Tissue Organ Culture*, **40**, 1–15.

Sonti, R.V., Chiurazzi, M., Wong, D., *et al.* (1995) *Arabidopsis* mutants deficient in T-DNA integration. *Proceedings of the National Academy of Sciences, USA*, **92**, 11786–11790.

Southgate, E.M., Davey, M.R., Power, J.B. & Marchant, R. (1995) Factors affecting the genetic engineering of plants by microprojectile bombardment. *Biotechnology Advances*, **13**, 631–651.

Stachel, S.E. & Nester, E.W. (1986) The genetic and transcriptional organization of the Vir region of the A6-Ti plasmid of *Agrobacterium tumefaciens*. *EMBO Journal*, **5**, 1445–1454.

Staub, J.M. & Maliga, P. (1992) Long regions of homologous DNA are incorporated into the tobacco plastid genome by transformation. *Plant Cell*, **4**, 39–45.

Staub, J.M. & Maliga, P. (1995) Expression of a chimeric uida gene indicates that polycistronic messenger-RNAs are efficiently translated in tobacco plastids. *Plant Journal*, **7**, 845–848.

Staub, J.M., Garcia, B., Graves, J., *et al.* (2000) High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nature Biotechnology*, **18**, 333–338.

Svab, Z., Harper, E.C., Jones, J.D.G. & Maliga, P. (1990) Aminoglycoside-3'-adenyltransferase confers resistance to spectinomycin and streptomycin in *Nicotiana tabacum*. *Plant Molecular Biology*, **14**, 197–205.

Svab, Z. & Maliga, P. (1993) High frequency plastid transformation in tobacco by selection for a chimeric aada gene. *Proceedings of the National Academy of Sciences, USA*, **90**, 913–917.

Tempe, J. & Casse-Delbart, F. (1989) Plant gene vectors and genetic transformation: *Agrobacterium* Ri plasmids. In: *Cell Culture and Somatic Cell Genetics of Plants, Volume 6* (ed. I.K. Vasil), pp. 25–49. Academic Press, London.

Tepfer, D. (1990) Genetic transformation using *Agrobacterium rhizogenes*. *Physiologia Planta*, **79**, 140–146.

Theologis, A., Zarembinski, T.I., Oeller, P.W., Liang, X.W. & Abel, S. (1992) Modification of fruit ripening by suppressing gene expression. *Plant Physiology*, **100**, 549–551.

Tiwari, V.K., Zhang, J., Golds, T.J., Cocking, E.C., Davey, M.R. & Power, J.B. (2001) Effect of heat shock treatment on *Hordeum vulgare* protoplast transformation mediated by polyethylene glycol. *Biologia Planta*, **44**, 25–31.

Touraev, A., Stoger, E., Voronin, V. & HeberleBors, E. (1997) Plant male germ line transformation. *Plant Journal*, **12**, 949–956.

Trail, F., Richards, C. & Wu, F.-S. (1989) Genetic manipulation in *Brassica*. In: *Biotechnology in Agriculture and Forestry, Volume 9. Plant Protoplasts and Genetic Engineering II* (ed. Y.P.S. Bajaj), pp. 197–215. Springer-Verlag, Berlin.

Trieu, A.T., Burleigh, S.H., Kardailsky, I.V., *et al.* (2000) Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant Journal*, **22**, 531–541.

Trifonova, A., Madsen, S. & Olesen, A. (2001) *Agrobacterium*-mediated transgene delivery and integration into barley under a range of in vitro culture conditions. *Plant Science*, **161**, 871–880.

Tzfira, T., Vaidya, M. & Citovsky, V. (2001) VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in VirE2 nuclear import and *Agrobacterium* infectivity. *EMBO Journal*, **20**, 3596–3607.

Tzfira, T., Vaidya, M. & Citovsky, V. (2002) Increasing plant susceptibility to *Agrobacterium* infection by overexpression of the *Arabidopsis* nuclear protein VIP1. *Proceedings of the National Academy of Sciences, USA*, **99**, 10435–10440.

van Attikum, H., Bundock, P. & Hooykaas, P.J.J. (2001) Non-homologous end-joining proteins are required for *Agrobacterium* T-DNA integration. *EMBO Journal*, **20**, 6550–6558.

Van Breusegem, F., Van Montagu, M. & Inze, D. (1998) Engineering stress tolerance in maize. *Outlook in Agriculture*, **27**, 115–124.

van Wordragen, M. & Dons, H. (1992) *Agrobacterium tumefaciens*-mediated transformation of recalcitrant crops. *Plant Molecular Biology Report*, **10**, 12–36.

Vandenelzen, P.J.M., Jongedijk, E., Melchers, L.S. & Cornelissen, B.J.C. (1993) Virus and fungal resistance – from laboratory to field. *Philosophical Transactions of the Royal Society, London Series B, Biological Sciences*, **342**, 271–278.

Villemont, E., Dubois, F., Sangwan, R.S., Vasseur, G., Bourgeois, Y. & Sangwan-Norreel, B.S. (1997) Role of the host cell cycle in the *Agrobacterium*-mediated genetic transformation of *Petunia*: evidence of an S-phase control mechanism for T-DNA transfer. *Planta*, **201**, 160–172.

Walden, R. (1988) *Genetic Transformation in Plants*. Open University Press, Milton Keynes.

Walden, R. & Wingender, R. (1995) Gene transfer and plant regeneration techniques. *Trends in Biotechnology*, **13**, 324–331.

Wang, H., Kraus, J., Dettendorfer, J., Chua, N.-H. & Nehls, R. (eds) (2002) Marker gene elimination from transgenic sugarbeet by chemically regulated Cre/Lox system. In: Plant Biotechnology 2002 and Beyond (I.K. Vasil, ed.), pp. 229–232. Proceedings of the 10th IAPTC&B Congress. Kluwer Academic Publishers, Orlando, Florida.

White, F.F. (1993) Vectors for gene transfer in higher plants. In: *Transgenic Plants, Volume 1* (eds S. Kung and R. Wu), pp. 15–48. Academic Press, San Diego.

Williams, M.E. (1995) Genetic engineering for pollination control. *Trends in Biotechnology*, **13**, 344–349.

Winans, S.C. (1991) An *Agrobacterium* 2-component regulatory system for the detection of chemicals released from plant wounds. *Molecular Microbiology*, **5**, 2345–2350.

Ye, G.N., Stone, D., Pang, S.Z., Creely, W., Gonzalez, K. & Hinchee, M. (1999) *Arabidopsis* ovule is the target for *Agrobacterium* in planta vacuum infiltration transformation. *Plant Journal*, **19**, 249–257.

Yoder, J.I. & Goldsbrough, A.P. (1994) Transformation systems for generating marker-free transgenic plants. *Bio-Technology*, **12**, 263–267.

Zaccomer, B., Cellier, F., Boyer, J.C., Haenni, A.L. & Tepfer, M. (1993) Transgenic plants that express genes including the 3' untranslated region of the Turnip Yellow Mosaic-Virus (TYMV) genome are partially protected against TYMV infection. *Gene*, **136**, 87–94.

Zambryski, P.C. (1992) Chronicles from the *Agrobacterium*-plant cell-DNA transfer story. Annual Reviews of Plant Physiology. *Plant Molecular Biology*, **43**, 465–490.

Zhan, X.C., Jones, D.A. & Kerr, A. (1988) Regeneration of Flax plants transformed by *Agrobacterium rhizogenes*. *Plant Molecular Biology*, **11**, 551–559.

Zhang, J., Boone, L., Kocz, R., Zhang, C.H., Binns, A.N. & Lynn, D.G. (2000) At the maize/*Agrobacterium* interface: natural factors limiting host transformation. *Chemistry and Biology*, **7**, 611–621.

Zhang, W. & and Wu, R. (1988) Efficient regeneration of transgenic plants from rice protoplasts and correctly regulated expression of the foreign gene in the plants. *Theoretical and Applied Genetics*, **76**, 835–840.

Zhong, H., Sun, B.L., Warkentin, D., *et al.* (1996) The competence of maize shoot meristems for integrative transformation and inherited expression of transgenes. *Plant Physiology*, **110**, 1097–1107.

Zhou, G., Weng, J., Zeng, Y., Huang, J., Qian, S. & Liu, G. (1983) Introduction of exogenous DNA into cotton embryos. *Methods in Enzymology*, **101**, 433–481.

Ziemienowicz, A., Gorlich, D., Lanka, E., Hohn, B. & Rossi, L. (1999) Import of DNA into mammalian nuclei by proteins originating from a plant pathogenic bacterium. *Proceedings of the National Academy of Sciences, USA*, **96**, 3729–3733.

Zupan, J., Muth, T.R., Draper, O. & Zambryski, P. (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. *Plant Journal*, **23**, 11–28.

Zupan, J. & Zambryski, P. (1997) The *Agrobacterium* DNA transfer complex. *Critical Reviews in Plant Science*, **16**, 279–295.

Zupan, J.R. & Zambryski, P. (1995) Transfer of T-DNA from *Agrobacterium* to the plant cell. *Plant Physiology*, **107**, 1041–1047.

5 Plant germplasm collections as sources of useful genes

Ian Godwin

5.1 Introduction

In *On the Origin of the Species*, Charles Darwin (1859) proposed that species are made up of variable populations, and that this variation is maintained by sexual reproduction. As any student of plant breeding knows, the basic premise of breeding activity is the availability of genetic variation. The practice of plant breeding – or, more generally, plant improvement – is harnessing and manipulating that genetic variation to reach a desirable outcome: the improved genotype or population of genotypes which make up a cultivated variety. Germplasm is the underlying genetic variation naturally available and, by extension, a germplasm collection is an attempt to sample the available germplasm and to conserve and characterize it – in other words, to make available as much genetic variation as possible.

5.1.1 Germplasm collections as museums

Collecting plant germplasm can be exciting, even hazardous work. There is an element of romance in trekking off to far-flung exotic locations, to brave hostile wild-life, the elements and political upheavals or even wars. Yet the adventures faced by the intrepid colonial European plant collectors and explorers on missions to introduce such staples as the snapdragon, poppy, Peonia rose and primrose to the European herbaceous border are a far cry from the modern real-life version of germplasm collectors and curators. Rather than simply stocking a royal botanic garden with a range of ornamentals, a germplasm collection needs to be a carefully planned and executed exercise.

Regardless of how extensive a germplasm collection is, it needs to be recognized that it represents a snapshot in time. To pretend that a germplasm collection is the preservation of the total genetic variation within a species, genus or family is naïve at best. The collection can only be regarded as a sample of the genetic variation taken at a particular time and geographic location.

5.1.2 The evolution and domestication of crop species

Many of the familiar species that provide our major nutrition are the result of millennia of domestication. In fact, four species – wheat, rice, maize and potato – provide approximately half of the world's human nutrition. A number of species only exist

as cultivated landraces and improved cultivars and are not found in the wild in their domesticated recognizable form. Important agricultural species have evolved and emerged from all continents.

Domestication has been the driving force in making crop plants more productive and generally of greater quality than their wild relatives. Important quality modifications such as the absence of toxic or unpalatable chemical compounds, bigger fruit, root or grain size, improved baking or oil qualities, and product colour variations have been important in delivering the diversity of foods available to many human consumers worldwide. In addition, domestication often involves changing perennial species into annual crops, and altered seed dispersal mechanisms. Seed dispersal traits such as brittle or shattering spikes in cereals like barley and wheat are actively selected against, and these species often become reliant on human intervention for reproduction and dispersal. A frequent outcome of domestication and the subsequent breeding and selection is the narrowing of the gene pool.

This phenomenon has been demonstrated with most of the world's major crop species. The recent application of molecular tools has revealed the extent of the evolutionary bottlenecks through which domestication has narrowed genetic diversity.

5.2 The concept of gene pools

5.2.1 The classical model

The gene pool model has been used to describe the genetic variation available when considering a particular species, with the primary gene pool seen as the major source of such variation (Figure 5.1) (Rick 1979). Harlan and deWet (1971) described a three gene pool model:

- *Primary*: most simply, the primary gene pool is the species, and consists of the genotypes which can intercross easily to produce fertile hybrids. It may include wild and weedy relatives of different subspecies, or perhaps even other related species.
- *Secondary*: this includes related species, usually within the same genus, which can intercross with the cultivated species and produce partially fertile, weak hybrids.
- *Tertiary*: this includes related species which will not normally or easily intercross. They may be made to produce hybrids with complicated techniques, such as bridging crosses and embryo rescue. The hybrids are generally weak with low fertility or sterility.

Historically, most plant breeding programmes have placed their emphasis on extensively mining the genetic variation within the species, or the primary gene pool. The obvious reasons for this are ease of crossing, similar mating systems, and fertility of the progenies. However, the concept of the species in flowering plants is not nearly

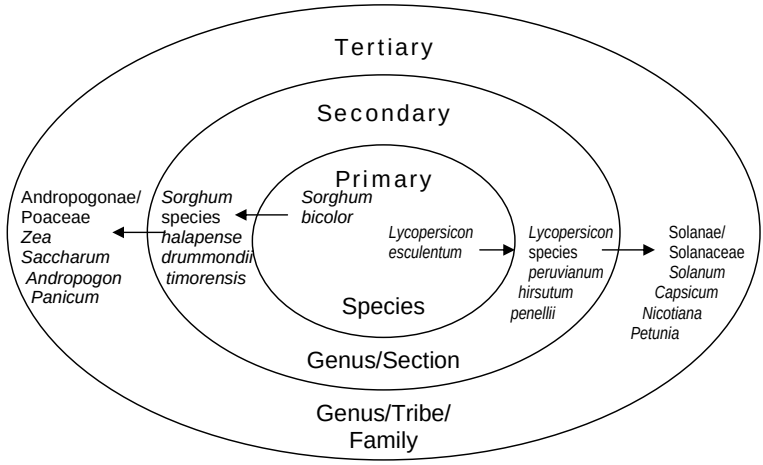


Figure 5.1 The gene pool model of the variation available for the improvement of a particular species, with sorghum and tomato as examples.

as clear-cut and absolute as that for mammals, for which the textbooks cite that two individuals are conspecific if they can mate and produce fertile offspring. Many plant species have fairly grey areas around the species boundary, and species such as sorghum and tomato will readily cross and produce fertile progenies with other 'species'. Cultivated grain sorghum, *Sorghum bicolor*, will readily hybridize with other members of the *Sorghum* section, including *S. verticilliflorum*, *S. propinquum* and *S. sudanensis*; hence, these are often regarded as subspecies of *S. bicolor*.

The familiar tomato and chilli (pepper) were found only in the Americas as recently as 500 years ago, yet today these vegetables are central to the cuisine of many European and Asian cuisines. *Lycopersicon esculentum* will readily hybridize with *L. pennellii* and *L. pimpinellifolium*. Indeed, the amount of genetic variation within the cultivated tomato is quite low. Restriction fragment length polymorphisms (RFLPs) among tomato cultivars are relatively infrequent (Van der Beek *et al.* 1992). As a result, many of the more useful genes used in breeding programmes have been introgressed from wild species within the genus. These include major genes for disease resistance, nematode resistance and fruit quality traits.

5.2.2 Beyond the tertiary gene pool

The transgenic era, which only began 20 years ago, has completely revolutionized the way in which we can view genetic resources. The total available genetic variation for any crop species lies not only within the family boundary, but beyond to include the entire plant kingdom. Hence, the variation existing within a simple weed can now be seen as a source of novel genes or alleles for our own favourite crop species. The boundaries can be pushed even further, and there is no doubt that all sources

of nucleic acid – whether plant, animal or microbial – are now potential donors of useful genetic variation. A quick glance at the transgenic crop plants currently in commercial production illustrates this perfectly, with maize and cotton containing insect resistance genes from the bacterium, *Bacillus thuringiensis*, and fruit and vegetable crops with virus resistance conferred by genes gleaned from the causal virus. Further to this, recombinant DNA technology offers new combinations of nucleic acid sequences not seen in natural systems, whereby a plant may be transformed with a chimeric gene originating from a range of species, as will be discussed further in Section 5.6.

5.3 Molecular genetics and genomics

Molecular markers, and particularly those based on DNA, have important roles to play in germplasm management. Avise (1994) stated that DNA markers are used most intelligently when they address areas of contention or are used to solve particular problems which are otherwise intractable using traditional techniques. Some strengths of these tools are:

1. Molecular data are genetic rather than phenotypic.
2. Molecular techniques are applicable to all organisms.
3. Molecular methods access a theoretically unlimited pool of genetic variability.
4. They can distinguish homology from analogy.
5. They provide a common measure for assessing divergence (Avise 1994).

There are numerous types of DNA markers available, with an even greater number of variants upon these with increasingly complex acronyms to describe them. All methods have characteristic advantages and disadvantages. Major considerations for the application of DNA markers in germplasm management and utilization include demonstrated ability of the system to deliver maximum informativeness in a cost-effective manner, and robust good laboratory-to-laboratory transferability.

DNA markers can be sub-divided into two major types: first, those relying on DNA hybridization to restriction-digested DNA; and second, those requiring the polymerase chain reaction (PCR). These marker systems are also discussed in Chapter 1.

5.3.1 RFLP markers

Recent studies have demonstrated the utility of RFLP analysis for genetic diversity analysis within species, such as sorghum (Tao *et al.* 1993). As a result of probe sequence homology, such markers are ideally suited to phylogenetic analysis between related species, such as within the genus *Oryza* (Wang *et al.* 1992), *Eleusine* (Salimath *et al.* 1995) and *Musa* (Bhat *et al.* 1997). RFLPs offer a very robust

methodology, with generally good transferability between laboratories. They are co-dominant, and hence can be good estimators of heterozygosity. Once a library of probes is generated, RFLPs can be generated without any sequence information.

However, there are considerable disadvantages with the technology, not least of which are the low levels of polymorphism seen within some species such as peanut (Kochert *et al.* 1991). For a previously unexplored species, such as taro (*Colocasia esculenta*), it is necessary to develop a suitable probe library, of either cDNA or genomic DNA. The generation of RFLP data is time-consuming, particularly with single-copy probes, and the assay is one of the most costly to perform as many steps are involved, usually with radioactive labels. Large quantities of DNA are also required, generally 5 to 10 μg per digest and, as a result, whole plants are often required for DNA extractions. Probes also need to be distributed to collaborating laboratories and, overall, the generation of RFLPs is moderately technically demanding.

5.3.2 PCR-based markers

PCR-based markers share a number of general advantages over RFLP technology. The major advantages are the speed with which results are generated, the low amounts of genomic DNA template required (5–50 ng, up to 1000-fold less than RFLP analysis), and the ability to share information on primer sequences without the need to exchange DNA.

These markers can be based on arbitrarily primed reactions, such as RAPDs (random amplified polymorphic DNA), ISSRs (inter simple sequence repeats) and AFLPs (amplified fragment length polymorphisms). However, many are also based on known sequences, such as microsatellites or SSRs (simple sequence repeats), STSs (sequence tagged sites) and SNPs (single nucleotide polymorphisms).

5.4 Plant germplasm collections

5.4.1 Collection issues: centres of diversity

Representative sampling of the genetic diversity of a species/genus requires knowledge of the centre of origin or centre of diversity for the particular species or genus in question. The Russian geneticist, Nikolai Vavilov, proposed eight 'centres of crop origin' – geographical areas which gave rise to much of the crop plant diversity and were centres of domestication (Table 5.1). These centres include the centres of origin for maize and sunflower from North America; potato, peanut and tomato from South America; rice, banana, soybean and taro from Asia; wheat and barley from the Middle East (the so-called 'fertile crescent'); cabbage and olives from Europe; and sorghum and millet from Africa. Australia hardly qualifies as a wild-food paradise, with the macadamia nut representing its only significant contribution as a continent to the global food supply.

Table 5.1 Vavilov's centres of crop origin. (After Vavilov 1926.)

Continent	Geographic region	Plant species
Asia	South-east Asia	Coconut, rice, sugarcane
	China	Chinese cabbages, soybean
	India	Cucumbers, eggplant, pigeonpea
	Turkey/Iran	Wheat, barley, oats, figs
North America	Mexico/Central America	Maize, tomato
South America	Andes/Brazil/Paraguay	Peppers, potato, rubber
Africa	Ethiopia	Sorghum, millets

5.4.1.1 Identification of phylogenetic relationships

Molecular techniques have had a great impact on the ability to identify phylogenetic relationships within genera, tribes and families. The diploid progenitors of allopolyploid species can be confirmed. The progenitor species of the sterile, parthenocarpic triploid and tetraploid bananas have been difficult to identify reliably, due largely to the inability to perform crosses with the parthenocarpic individuals. DNA markers have been used to demonstrate that the chloroplast genome of bananas is inherited maternally, while the mitochondrial genome is paternally inherited (Faure *et al.* 1993). This knowledge has allowed the elucidation of the origin of cultivars and the relationships among the many diploid species and subspecies of the *Musa* genus. The wild progenitors of cultivated species may be reliably identified. Markers have been used to identify the wild progenitors of diverse species such as cultivated maize (Doebley 1989) and finger millet (Salimath *et al.* 1995).

The classic study of Bonierbale *et al.* (1988) revealed the extent of similarity between tomato and potato in genome organization. Colinearity or synteny was also identified between sorghum and maize (Berhan *et al.* 1993) and subsequently extended to all major grass genomes. These relationships are not merely academically important in allowing questions to be answered regarding crop species evolution, but the colinearity can greatly expedite the cloning of genes or marker-assisted selection for genes for traits in related species (Gepts 1995). The availability of information from genome-sequencing programmes such as the rice and *Arabidopsis* genome projects have meant that colinearity can facilitate more rapid cloning of genes from related species. The rice genome sequence is useful for all plant improvement programmes, but specifically the other members of the family Poaceae, the grasses and cereals. Knowledge of the map-position of a particular locus for a trait in rice, will expedite the identification of DNA markers linked to the homologue in wheat, barley, oats, sorghum, maize and sugarcane (see Chapter 3).

5.4.2 Management and conservation: rationalization and core collections

A successful germplasm collection will contain materials for short-term benefit and long-term insurance (Kresovich & McFerson 1992). One of the greatest challenges facing the curator of a germplasm question is the efficiency and costs of storage, re-

generation and maintenance of genetic integrity whilst conserving as much genetic diversity as possible. Molecular tools are extremely useful in the management and conservation of germplasm collections.

DNA fingerprint data can be applied to answer questions of:

- identity;
- duplication;
- genetic diversity;
- contamination; and
- integrity of regeneration.

In addition, for species which are vegetatively propagated such as potato, sugarcane, taro and sweet potato, DNA markers are extremely powerful for identifying zygosity at important loci. DNA markers may also be used, in conjunction with other data, to select core collections, and this should facilitate the more efficient utilization of genetic resources for plant breeding and population enhancement. Recent application of these techniques has been made to collect and manage taro (*Colocasia esculenta*) genetic resources in the south Pacific Island countries.

The genetic diversity of taro genetic resources has been previously characterized largely by morphological and cytological variation (Kuruvialla & Singh 1981; Tanimoto & Matsumoto 1986), and it has been observed that Polynesian cultivars are highly morphologically variable in contrast to the phenotypic homogeneity of the wild populations of Melanesia. It is thought that the high level of phenotypic variation is due to a high rate of vegetative propagation and, consequently, of somatic mutations. This would suggest that the majority of the cultivars in Polynesia are clones of a common source, and a recent study using isoenzymes (Lebot & Aradhya 1991) indicated that there was very little genetic variation between the Polynesian cultivars, in contrast to the Melanesian and Asian cultivars. The results from a molecular study of taro genetic diversity, using RAPDs (Irwin *et al.* 1998), confirmed that although the cultivars in the Pacific region exhibit remarkable morphological variation, the genetic base appears to be very narrow. Such a limited genetic base leaves the crop very vulnerable to disease epidemics and insect damage.

Consequently, germplasm collections from around the region have been undertaken to augment existing national collections and to safeguard threatened and useful germplasm for use in regional breeding programmes. Some 1500 accessions are currently recognized by the Taro Genetic Resources Network (TaroGen), which aims to establish a regional genebank with a core collection, representative of the genetic diversity found within all the national collections. The germplasm must be well-characterized for the establishment of such a collection, in addition to the subsequent management and utilization of the collection, in order to minimize genotypic redundancy, identify gaps in the collections and hence maximize the genetic diversity conserved. This is particularly important in germplasm collections of clonally propagated species, which frequently contain accessions which, although phenotypically similar, have different genetic origins. Furthermore, identi-

cal cultivars may have different names throughout the region, and this is especially true in Melanesia where numerous vernacular languages increase the chances of synonymy and duplication (Lebot & Aradhya 1991). To date, studies directed at the identification of redundant germplasm in the Pacific Island national collections have utilized biogeographic, agronomic and phenotypic characterization. Increasingly, the characterization of germplasm collections also utilizes molecular techniques (Teulat *et al.* 2000; van Treuren *et al.* 2001). The availability of PCR-based markers, such as SSRs, offers the opportunity for more fine-scale genetic characterization of germplasm collections than previously possible, due to their high levels of polymorphism, their occurrence throughout the genome, their ease of detection, and the additional advantage that many of the complications of environmental effects acting upon characters is avoided by looking directly at variation controlled at the genetic level. SSRs are one of the markers of choice for many plant breeding applications, due particularly to their co-dominant nature, transferability, reproducibility and amenability to high throughput.

SSR markers are increasingly used for investigations of critical importance to curators of germplasm collections, including:

1. The establishment of unique genetic identities or fingerprints (Hokanson *et al.* 1998; Becher *et al.* 2000; Prasad *et al.* 2000).
2. The assessment of genetic diversity contained within a collection (Westman & Kresovich 1999; Macaulay *et al.*, 2001).
3. The determination of genetic relatedness between accessions (e.g. Lopes *et al.* 1999; Li *et al.* 2000).
4. The evaluation of core collections (e.g. Gepts 1995; Huaman *et al.* 2000), where a core collection of plant genetic resources consists of a limited set of accessions chosen to represent the genetic variation in a crop species and its wild relatives with the minimum of repetition (Brown & Spillane 1999).

Traditionally, geographical (passport) data and morphological characterization data have been used for the development of core collections (Ford-Lloyd 2001). Microsatellite markers have previously been isolated from taro (Mace & Godwin 2002) and a set of polymorphic markers identified through screening with a limited range of genotypes from the Pacific Island region. These markers have been applied to select a core collection which represents 10% of the total collection, but greater than 70% of the available genetic diversity.

5.4.3 Utilization requires some sort of screening

It must be acknowledged that as well as representing only a snapshot in time for any given species, a germplasm collection has no intrinsic value. A collection of crop genotypes is not like a collection of Italian masters' artworks. Its raison d'être lies in its utilization, and in general a germplasm collection has little to recommend it from the viewpoint of beauty. This may be disputed for a species such as sunflower,

where the diversity of flower shape and colour when growing in a field genebank could be argued to share the beauty of a Van Gogh or Monet original. It could also be argued that the value of such a collection may actually surpass that of Van Gogh's still-life painting 'Sunflowers'. On the last occasion that this painting was auctioned it sold for the astronomical sum of $39 million. The major disease of sunflower worldwide is sclerotinia wilt, caused by the fungus, *Sclerotinia sclerotorium*. In the United States alone, the disease causes a loss of $17.2 million in production annually, with another $4.8 million lost to changes in production costs (Gianessi *et al.* 2002). Hence, a single gene conferring resistance – even for as little as a few years – would far surpass even the most expensive artwork in 'value' on current figures. Public germplasm collections and the free distribution of genetic variation and information for manipulation in plant breeding are very valuable resources indeed. In recent attempts to measure the value of germplasm and its manipulation, Byerlee and Traxler (1995) estimated that the annual benefits of breeding for spring wheat in the developing world was around US$2.5 billion, More recently, Cassady *et al.* (2001) estimated that total outcomes of wheat breeding internationally may be as valuable as US$4 billion annually in increased production efficiency and reduced costs associated with biotic and abiotic stresses.

5.5 The challenge for plant breeding: utilization

As already stated, the net worth of a germplasm collection lies in its utilization in crop improvement. The greatest difficulty in measuring the use of a single accession is that whether it will be used, and when it may be used, is commonly unforeseeable. For species which exist in the wild (such as *Sorghum bicolor*) it must be accepted that in a majority of cases, many accessions will never donate a single gene to an improved cultivar. However, in some cases, the species does not exist in an undomesticated form. Hexaploid wheat (*Triticum aestivum*) is an example, and as the most collected of all species (Table 5.2), it could be surmised that many of accessions in the collection were at one stage contributing to human food intake. Note however, that most wheat collections will contain not only cultivated hexaploid bread wheat, but also tetraploid and diploid progenitors and wild relatives from the secondary and tertiary gene pools. These will include not only *Triticum* spp., but related genera such as *Aegilops* and *Agropyron*, and perhaps other cultivated genera such as *Secale.*

Utilization relies on the recognition of worth of a particular accession, or perhaps even a particular gene or allele conferring a sought-after trait. Such a trait may be disease or insect resistance, a particular quality attribute such as baking quality, digestibility or seed coat colour, or perhaps a physiological component of yield such as height, seed size or water-use efficiency. If the trait is a rare allele, then finding an accession with such a trait may be extremely difficult. Complete phenotyping of a complete collection is usually an economic and temporal impossibility. Imagine the time and effort need to phenotype even a single collection of moderate size!

Table 5.2 Germplasm collections of major crop species based on number of accessions held in *ex situ* collections worldwide. (After FAO 1996.)

Crop type	Species	Number of accessions
Cereals	Wheat	784 500
	Barley	485 000
	Rice	420 500
	Maize	277 000
	Oat	222 500
	Sorghum	168 500
Pulse/Legume	*Phaseolus* bean	268 500
(including oilseed)	Soybean	174 500
	Cowpea	85 500
	Peanut/Groundnut	81 000
Vegetable	*Brassica*	109 000
	Tomato	78 000
	Capsicum	53 500
Fruit	Apple	97 500
	Prunus	64 500
	Grape	47 000

The ICRISAT sorghum collection consists of approximately 36 700 accessions. The task of phenotyping such as collection for a small number of traits such as height, phenology, biomass, seed colour and size and plant colour is not a trivial undertaking. Hence, to then consider searching for a potentially rare allele (ergot resistance, *Striga* resistance, high amylose endosperm), or a difficult/expensive to measure trait (kafirin protein classes, water-use efficiency, osmotic adjustment) means that such data will never be available for more than just a fraction of the collection.

5.5.1 Exploiting the primary gene pool

The cultivated members of the species *Beta vulgaris* include sugarbeet, garden and leaf beets, and fodder beets. Wild relatives within the primary gene pool include the subspecies *maritima* and have been used for decades to improve traits such as sugar content (Tjebbes 1933) and *Cercospora* leaf spot resistance (Munerati 1932). Accessions of this sub-species have been shown to carry sources of resistance to a range of biotic stresses, including viral, fungal and nematode pathogens (summarized in Frese *et al*. 2002).

5.5.2 Exploiting the secondary gene pool

The secondary gene pool of tomato includes many species of the *Lycopersicon* genus, and this genus has been a major contributor of agronomic and quality traits to the modern cultivated tomato. Single genes conferring traits such as cucumber

mosaic virus resistance from *L. chilense* (Stamova & Chetelet 2002) and the sucrose accumulator gene for quality from *L. chmielewskii* (Chetelet *et el.* 1995) have been introgressed. More complex traits or quantitative trait loci (QTL) such as black-mould resistance from *L. cheesmanii* (Robert *et al.* 2001) and fruit characters and genes involved in heterosis from *L. hirsutum* (Monforte & Tanksley 2000) have also been introgressed with the aid of DNA marker-assisted selection.

The genus has been well studied taxonomically, and the applications of isoenzymes and DNA markers have been effective in evaluating genetic diversity and species relationships within the genus. Isozymes have been generally too few to be useful for these purposes (Breto *et al.* 1993). Co-dominant markers such as RFLPs (Miller & Tanksley 1990) and microsatellites (Alvarez *et al.* 2001) have been demonstrated to be informative for phylogenetic studies of tomato and its relatives within *Lycopersicon*. Microsatellite markers are often the more informative of these two marker types, yet often the difficulties associated with cloning and sequencing microsatellites is seen as a great impediment to their application for germplasm analysis. Multilocus dominant markers have also been successfully applied to understand and quantify genetic diversity and species relationships within genera. These include markers such as RAPDs in sweet potato (Connolly *et al.* 1994) and rice (Virk *et al.* 1995), AFLPs in eggplant and wild relatives (Mace *et al.* 1999) and *Lactuca* (Hill *et al.* 1996) and ISSRs in *Eleusine* (Salimath *et al.* 1995) and sorghum (Yang *et al.* 1996).

5.5.3 Beyond the secondary gene pool: how can barriers be broken down?

The greatest barrier to utilization of the tertiary gene pool is the difficulty associated with introgression. The successful transfer of a trait from a related species/genus to the domesticated species will often require the use of a technique such as embryo rescue or the use of bridging species for crosses, or in some cases asexual means such as protoplast fusion – as has been widely used in citrus improvement.

A number of techniques to facilitate interspecific/intergeneric hybridizations have been successfully used in many crop species. These include the use of irradiation or chemical mutagens to facilitate crossing, and often the use of alien addition or substitution lines. Somatic protoplast fusions have been used to form interspecific hybrid plants between species such as potato and tomato (Melchers *et al.* 1978).

5.5.3.1 Diploid progenitors

Diploid progenitor species have been widely used during recent decades to introgress genes into allopolyploid species. In many cases, the use of diploid progenitors has been extended to resynthesize the species by crossing, embryo rescue and chromosome doubling using colchicine. Synthetic polyploid plants are extremely useful for the introgression of novel genes for important traits, as they seek to overcome the evolutionary bottleneck which has arisen as the result of a hybridization thought to have occurred in recent (post-agricultural) history.

Wheat (*Triticum aestivum*) has been resynthesized by crosses between the tetraploid AABB genome *Triticum turgidum* (durum wheat) and the diploid DD genome *Triticum tauschii* (syn. *Aegilops tauschii,* goat grass). This has been performed in a number of institutions, worldwide, and has been a major output of the CIMMYT wheat-breeding programme over the past decade. *T. tauschii* is the source of many stress resistances which are not found within cultivated wheats (Appels & Lagudah 1990; May & Lagudah 1992), and a number of resistance gene traits have now been mapped to the D genome in synthetic wheats (Arraiano *et al.* 2001).

Tetraploid canola, or oilseed rape (*Brassica napus*) has been successfully resynthesized from its diploid progenitors, *B. campestris* and *B. oleracea* (Song *et al.* 1993) using similar techniques. Synthetic *Brassicas* have been used to generate new variant genotypes for phenology, and these show promise in the further breeding of canola. Synthetic allotetraploids have also been used to study the effects of polyploidization on genome alterations (Song *et al.* 1995).

5.6 Modern techniques: genomics meets bioinformatics

The utilization and management of genetic resources have been revolutionized by major advances in molecular biology and information technology. Molecular biology has allowed the use of DNA markers and DNA sequencing for better description and manipulation of genetic diversity, and the cloning and transgenic manipulation of traits. Information technology has allowed for an explosion of gene sequence and trait data and the international access to databases through the internet. Some of the ways in which molecular techniques and information technology have led to improved management and utilization of genetic resources are described below.

5.6.1 Better description and measurement of diversity

As already discussed, molecular genetics and genomics techniques can provide great insights into genome structure, evolution and the amount of genetic diversity within a species. One of the challenges to be faced for germplasm collectors and curators is where to collect diversity from most efficiently. The ability to adequately analyse eco-geographic patterns in germplasm diversity can assist with planning collecting activities, targeting specific variation for utilization and developing core collections. One such tool is the DIVA-GIS software, developed by CIP, the International Potato Centre and IPGRI (Hijmans *et al.* 2001). This integrates genetic diversity information, such as DNA marker data, with geographic information system (GIS) and mapping software. It allows the mapping of genetic resource richness and diversity, distributions of desirable markers or traits and location of areas with complementary diversity.

5.6.2 Better screening of variation for a trait

DNA markers have been proposed as a means of screening genetic resources for particular alleles for valuable traits. Virk *et al.* (1996) used RAPD markers as a means of identifying variants for a number of quantitative traits within rice germplasm. Association mapping also has potential to identify markers associated with quantitative traits (Beer *et al.* 1997).

5.6.3 Better management of data: dynamic data sets

DNA fingerprint data may be stored as binary code, and as scanned images of gels and autoradiographs to facilitate the identification of accessions, and for the analysis of genetic diversity and phylogeny. This represents the simplest form of storing DNA data, and it can be done as part of a germplasm collection database where information is stored with passport, morphological and agronomic trait data.

This can then be made more complex, and more useful by the inclusion of molecular map data, DNA and protein sequence data, and data on quantitative traits. These can be made rapidly and publicly available via the internet or World Wide Web. Increasingly, DNA and germplasm databases are being integrated in this manner. One of the first to become available was the GrainGenes database, administered by the USDA. This database concentrated on the Triticeae, with particular emphasis on: comparative mapping between species, gene and allele catalogues, genotypes of cultivars and breeding lines regarding disease resistance, quality and agronomic traits, taxonomy and disease and pest information (Matthews & Anderson 1997).

Through integration with major DNA sequence and genome websites, including GenBank – which is accessible through many mechanisms, including the US National Centre for Biotechnology Information (www.ncbi.nlm.nih.org) – the ability to compare and identify sequences for conserved and divergent motifs and functions has become widely available. In concert with germplasm databases, these can be made into very powerful tools for genetic resource management and utilization. One such development is the International Crop Information System, or ICIS. This is a collaborative venture to develop database systems for the management and integration of global information on genetic resources, and is aimed at being useful for any crop species. It involves scientists from a number of international institutions, particularly those centres from the CGIAR network, such as CIMMYT, ICRISAT and IRRI. The aim is to provide an information system for each individual crop, and will enable the integration of data from different crops using the Systemwide Information Network for Genetic Resources (SINGER). Collaborative projects will enable dynamic databases to be developed, and will accept real-time interrogation and data entry from network members. This will lead to better access to data and sharing of information by international collaborative groups. CIMMYT for example, have developed the International Wheat Information System (IWIS) based on ICIS, which manages information pertaining to bread and durum wheat and triticale (www.cimmyt.org). IWIS consists of two major components: the Pedigree Manage-

ment System, which assigns unique genotype identifier and pedigrees; and the Data Management System, which is a database of field and laboratory studies, and data on known genes. IWIS is accessible to registered users via the World Wide Web, and is distributed as a CD-ROM on a regular basis when updates become available.

5.7 Beyond the gene pool: redesigning agricultural species with germplasm utilization

Molecular biology, genomics and information technology have led to a bright future for the utilization of genetic resources for crop plant improvement. The availability of whole genome sequences will allow functional analysis of genes in new backgrounds and the molecular dissection of complex traits. These new technologies will lead to exciting new opportunities for germplasm utilization and enhancement. The gene pool for any given species has expanded well beyond the tertiary gene pool, and can be taken to include any gene from any source, perhaps even new synthetic or 'shuffled' sequences (Table 5.3).

There are nevertheless, some reasons for caution. The applications of these new technologies must be made both cleverly and with sensitivities to the wants and needs of societies in both the developed and developing world. There are complex scientific, biosafety and ethical considerations to be addressed with the application of these technologies. Most societies want to see the benefits to the 'triple bottom line' of economy, environment and society. There are also complex social and political sensitivities associated with the collection and utilization of genetic resources from developing regions of the world, and questions about the rights to intellectual property in centres of diversity of crop species where the major benefits of improvement may be mainly realized in the developed nations. There are therefore requirements for international regulations and protocols governing the collection, distribution and

Table 5.3 Origin of recombinant DNA in genetically modified (GM) crop plants: genetic resources beyond the tertiary gene pool.

GM crop plant	Trait gene origins	Marker gene origin
Roundup Ready soybean	Bacterial Plant	Bacterial
Bt corn for corn borer resistance	Bacterial: *Bacillus thuringiensis*	Bacterial
Golden rice (provitamin A)	Plant: *Narcissus* Bacterial: *Erwinia*	Bacterial
Papaya ringspot virus-resistant papaya	Viral: PRSV	Bacterial
FlavrSavr tomato	Plant: *Lycopersicon*	Bacterial

utilization of genetic resources which go beyond the mere technologies of achieving plant genetic improvement, and scientists need to ensure that they are part of the total process to deliver public benefit where it is most needed.

References

Alvarez, A.E., van de Wiel, C.C.M,, Smulders, M.J.M. & Vosman, B. (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. *Theoretical and Applied Genetics*, **103**, 1283–1292.

Appels, R. & Lagudah, E.S. (1990) Manipulation of chromosomal segments from wild wheat for the improvement of bread wheat. *Australian Journal of Plant Physiology*, **17**, 253–266.

Arraiano, L.S., Worland, A.J., Ellerbrook, C. & Brown, J.K.M. (2001) Chromosomal location of a gene for resistance to *Septoria tritica* blotch (*Mycosphaerella graminicola*) in the hexaploid wheat 'Synthetic 6X'. *Theoretical and Applied Genetics*, **103**, 758–764.

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Becher, S.A., Steinmetz, K., Weising, K., *et al*. (2000) Microsatellites for cultivar identification in *Pelargonium*. *Theoretical and Applied Genetics*, **101**, 643–651.

Beer, S.C., Siripoonwiwat, W., O'Donoughue, L.S., Sousza, E., Matthews, D. & Sorrells, M.E. (1997) Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages? *Journal of Agricultural Genomics* 3 (http://www.ncgr.org/research/jag/papers97/paper197/index197.html).

Berhan, A.M., Hulbert, S.H., Butler, L.G. & Bennetzen, J.L. (1993) Structure and evolution of the genomes of *Sorghum bicolour* and *Zea mays*. *Theoretical and Applied Genetics*, **86**, 598–604.

Bhat, K.V., Lakhanpaul, S., Chandel, K.P.S. & Jarret, R.L. (1997) Molecular markers for characterization and classification of genetic resources of perennial crops. In: *Molecular Genetic Techniques for Plant Genetic Resources* (eds W.G. Ayad, T. Hodgkin, A. Jaradat & V.R. Rao), pp. 107–117. IPGRI, Rome, Italy.

Bonierbale, M.W., Plaisted, R.L. & Tanksley, S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, **120**, 1095–1103.

Breto, M.P., Asins, M.J. & Carbonell, E.A. (1993) Genetic variability in *Lycopersicon* species and their genetic relationships. *Theoretical and Applied Genetics*, **86**, 113–120.

Brown, A.H.D. & Spillane, C. (1999) Implementing core collections – principles, procedures, progress, problems and promise. In: *Core Collections for Today and Tomorrow* (eds R.C. Johnson & T. Hodgkins), pp. 1–9. International Plant Genetic Resources Institute, Rome, Italy.

Byerlee, D. & Traxler, G. (1995) National and international wheat improvement research in the post-green revolution period – evolution and impacts. *American Journal of Agricultural Economics*, **77**, 268–278.

Cassaday, K., Smale, M., Fowler, C. & Heisey, P.W. (2001) Benefits from giving and receiving genetic resources: the case of wheat. *Plant Genetic Resources Newsletter*, **127**, 1–10.

Connolly, A.G., Godwin, I.D., Cooper, M. & DeLacy, I.H. (1994) Interpretation of randomly amplified polymorphic DNA data for fingerprinting sweet potato. *Theoretical and Applied Genetics*, **88**, 332–336.

Chetelet, R.T., DeVerna, J.W.,& Bennett, A.B. (1995) Effects of *Lycopersicon chmielewskii* sucrose accumulator gene (*sucr*) on fruit yield and quality parameters following introgression into tomato. *Theoretical and Applied Genetics*, **91**, 334–339.

Darwin, C. (1859) *On the Origin of the Species by Means of Natural Selection*. John Murray, London.

Doebley, J. (1989) Isozymic evidence and the evolution of crop plants. In: *Isozymes in Plant Biology* (eds D. Soltis & P. Soltis), pp. 165–191. Dioscorides, Portland.

FAO (1996) *Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. FAO, Rome, 75pp.

Faure, S., Noyer, J.L., Carreel, F., Horry, J.P., Bakry, F. & Lanaud, C. (1994). Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Current Genetics*, **25**, 265–269.

Ford-Lloyd, B.V. (2001) Genotyping in plant genetic resources. In: *Plant Genotyping: The DNA Fingerprinting of Plants* (ed. R.J. Henry), pp. 59–82. CAB International, UK.

Frese, L. (2002) Combining static and dynamic management of PGR: a case study of Beta genetic resources. In: *Managing Plant Genetic Diversity* (eds J.M.M. Engels, V. Ramanatha Rao, A.H.D. Brown, & M.T. Jackson), Proceedings of an International Conference, Kuala-Lumpur, Malaysia,12–16 June 2000, pp. 133–147.

Gepts, P. (1995). Genetic markers and core collections. In: *Core Collections of Plant Genetic Resources* (eds T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum & E.A.V. Morales), pp. 127–146. John Wiley & Sons, Chichester, UK.

Gianessi, L.P., Silvers, C.S., Sankula, S. & Carpenter, J. (2002) *Plant Biotechnology: Current and Potential Impact for Improving Pest Management in U.S. Agriculture*. National Centre for Food and Agricultural Policy.

Harlan, J.R. & deWet, J.M.J. (1971) Towards a rational classification of cultivated plants. *Taxon*, **20**, 509–517.

Hill, M., Witsenboer, H., Zabeau, M., Vos, P., Kesseli, R. & Michelmore, R. (1996) PCR based fingerprinting using AFLPs as a tool for studying genetic relationships in the *Lactuca* spp. *Theoretical and Applied Genetics*, **93**, 1202–1210.

Hijmans, R.J., Guarino, L., Cruz, M. & Rojas, E. (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter*, **127**, 15–19.

Hokanson, S.C., Szewc-McFadden, A.K., Lamboy, W.F. & McFerson, J.R. (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* × *domestica* borkh. core subset collection. *Theoretical and Applied Genetics*, **97**, 671–683.

Huaman, Z., Ortiz, R., Zhang, D. & Rodriguez, F. (2000) Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp *andigena* potato cultivars. *Crop Science*, **40**, 273–276.

Irwin, S.V., Kaufusi, P., Banks, K., de la Peña, R. & Cho, J.J. (1998) Molecular characterisation of taro (*Colocasia esculenta*) using RAPD markers. *Euphytica*, **99**, 183–189.

Kochert, G., Halward, T., Branch, W.D. & Simpson, C.E. (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theoretical and Applied Genetics*, **81**, 565–570.

Kresovich, S. & McFerson, J.R. (1992) Assessment and management of plant genetic diversity – considerations of intraspecific and interspecific variation. *Field Crop Research*, **29**, 185–204.

Kuruvialla, K.M. & Singh, A. (1981) Karyotypic and electrophoretic studies on taro and its origins. *Euphytica*, **30**, 405–412.

Lebot, V. (1992) Genetic vulnerability of Oceania's traditional crops. *Experiments in Agriculture*, **29**, 309–323.

Lebot, V. & Aradhya, K.M. (1991) Isozyme variation in taro (*Colocasia esculenta* (L.) Schott.) in Asia and Oceania. *Euphytica*, **56**, 55–66.

Li, C.D., Rossnagel, B.G. & Scoles, G.J. (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. *Theoretical and Applied Genetics*, **101**, 1259–1268.

Lopes, M.S., Sefc, K.M., Eiras Dias, E., Steinkellner, H., Laimer da Câmara Machado, M. & da Câmara Machado, A. (1999) The use of microsatellites for germplasm management in a Portuguese grapevine collection. *Theoretical and Applied Genetics*, **99**, 733–739.

Macaulay, M., Ramsay, L., Powell, W. & Waugh, R. (2001) A representative, highly informative 'genotyping set' of barley SSRs. *Theoretical and Applied Genetics*, **102**, 801–809.

Mace, E.S. & Godwin, I.D. (2002) Development and characterisation of polymorphic microsatellite markers in taro, *Colocasia esculenta* (L.) Schott. *Genome*, **45**, 823–832.

Mace, E.S., Lester, R.N. & Gebhart, C.G. (1999). AFLP analysis of genetic relationships among cultivated eggplant, *Solanum melongena* and its wild relatives (*Solanaceae*). *Theoretical and Applied Genetics*, **99**, 626–633.

Matthews, D.E. & Anderson, O.D. (1997) Managing, storing and using molecular data. In: *Molecular Genetic Techniques for Plant Genetic Resources* (eds W.G. Ayad, T. Hodgkin, A. Jaradat & V.R. Rao). IPGRI, pp. 82–89.

May, C.E. & Lagudah, E.S. (1992) Inheritance in hexaploid wheat of *Septoria tritici* blotch resistance and other characteristics derived from *Triticum tauschii*. *Australian Journal of Agricultural Research*, **43**, 433–442.

Miller, J.C. & Tanksley, S.D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theoretical and Applied Genetics*, **80**, 437–448.

Melchers, G., Sacristan, M.D. & Holder, A.A. (1978) Somatic hybrid plants of potato and tomato regenerated from fused protoplasts. *Carlsberg Research Communications*, **43**, 203–218.

Monforte, A.J. & Tanksley, S.D. (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theoretical and Applied Genetics*, **100**, 471–479.

Munerati, O. (1932) Sull' incrocio della barbebietola coltivata con la beta selvaggia della costa adriatica. *L'Industria Saccarifera Italiana*, **25**, 303–304.

Prasad, M., Varshney, R.K., Roy, J.K., Balyan, H.S. & Gupta, P.K. (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. *Theoretical and Applied Genetics*, **100**, 584–592.

Rick, C.M. (1979) Biosystemic studies in *Lycopersicon* and closely related species of *Solanum*. In: *The Biology and Taxonomy of the Solanaceae* (eds J.G. Hawkes, R.N. Lester & A.D. Skelding), pp. 667–678. Linnean Society Symposium Series 7, Academic Press, London.

Robert, V.J.M., West, M.A.L., Inai, S., *et al.* (2001) Marker-assisted introgression of blackmold resistance QTL alleles from wild *Lycopersicon cheesmanii* to cultivated tomato (*L. esculentum*) and evaluation of QTL phenotypic effects. *Molecular Breeding*, **8**, 217–233.

Salimath, S.S., de Oliveira, A.C., Godwin, I.D. & Bennetzen, J.L. (1995) Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome*, **38**, 757–763.

Song, K., Tang, K. & Osborn, T.C. (1993) Development of synthetic *Brassica* amphidiploids by reciprocal hybridization and comparison to natural amphidiploids. *Theoretical and Applied Genetics*, **86**, 811–821.

Song, K.M., Lu, P., Tang, K.L. & Osborn, T.C. (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences, USA*, **92**, 7719–7723.

Stamova, B.S. & Chetelet, R.T. (2000) Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from *Lycopersicon chilense* into tomato. *Theoretical and Applied Genetics*, **101**, 527–537.

Tanimoto, T. & Matsumoto, T. (1986) Variations of morphological characters and isozyme patterns in Japanese cultivars of *Colocasia esculenta* Schott and *C. gigantea* Hook. *Japanese Journal of Breeding*, **36**, 100–111.

Tao, Y., Manners, J.M., Ludlow, M.M. & Henzell, R.G. (1993). DNA polymorphisms in grain sorghum (*Sorghum bicolor* (L.) *Moench*). *Theoretical and Applied Genetics*, **86**, 679–688.

Teulat, B., Aldam, C., Trehin, R., *et al.* (2000) An analysis of genetic diversity in coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. *Theoretical and Applied Genetics*, **100**, 764–771.

Tjebbes, K. (1933) The wild beets of the North Sea region. *Botaniska Notiser*, **14**, 305–315.

Van der Beek, H., Verkerk, R., Zabel, P. & Lindout, P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics*, **84**, 106–112.

van Treuren, R., van Soest, L.J.M. & van Hintum, T.J.L. (2001) Marker-assisted rationalisation of genetic resource collections: a case study in flax using AFLPs. *Theoretical and Applied Genetics*, **103**, 144–152.

Virk, P.S., Ford-Lloyd, B.V., Jackson, M.T., Pooni, H.S., Clemeno, T.P. & Newbury, H.J. (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, **76**, 296–304.

Wang, Z.Y., Second, G. & Tanksley, S.D. (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theoretical and Applied Genetics*, **83**, 565–581.

Westman, A.L. & Kresovich, S. (1999) Simple sequence repeat (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. *Euphytica*, **109**, 85–92.

Yang, W., de Oliveira, A.C., Godwin, I., Schertz, K. & Bennetzen, J.L. (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: high levels of variability observed in Chinese sorghum. *Crop Science*, **36**, 1669–1676.

Vavilov, N.I. (1926) Studies on the origin of cultivated plants. *Bulletin of Applied Botany, Genetics and Plant Breeding*, **16**, 3–248 (in Russian).

6 The impact of plant genomics on maize improvement

Donal M. O'Sullivan and Keith J. Edwards

6.1 Introduction

Maize is unique among the cereal crops, both in the geographical extent of its cultivation and in its diversity of end uses. Thousands of years of adaptation and selection have led to the creation of a multitude of varieties of maize that suit the climates in which they are grown and serve the needs of the people who grow them. This remarkable phenotypic diversity of cultivated maize is mirrored by the diversity of its genetic make-up. It is clear therefore that breeders have an immense resource available to them in the task of tailoring maize – which is still an important source of food and wealth – to an ever-changing set of economic and environmental parameters.

There are many reasons why genomics will play a critical role in future maize improvement strategies. First, the complexity of the maize genome (estimated at 2500 Mb and containing up to 50 000 genes) requires high-throughput approaches to assist in the location and characterization of genes of interest. Second, most traits to be manipulated will be genetically complex, involving inputs from many genes that participate in or regulate the pathways involved. In order to understand these complex interactions, the ability to monitor simultaneously the expression levels of each component gene product is essential. Third, each maize gene may occur in a virtually infinite range of variations. These variations must be recorded on a large scale in order to understand the basis for phenotypic variation.

In this chapter, we present maize in the context of its considerable diversity of form and function, and also present an overview of targets and methodologies important to the current generation of maize breeders. The review of past progress in maize improvement – which we have labelled 'Twentieth century maize breeding' – includes consideration of advances in maize molecular genetics which open up new possibilities for maize improvement, even though their impact has yet to reach the field. Looking forward to an even more far-reaching revolution in our understanding of the molecules of which maize is constructed, we review the status of genomics platforms being developed for maize and other plants, we examine how they may be integrated, and we discuss the impact that these may have on maize breeding. We conclude by bringing together a vision for the potential for 'molecular breeding' of maize.

6.1.1 Origins and distribution of maize growing

It is now widely accepted that cultivated maize (*Zea mays* L. spp. *mays*) was domesticated from its wild grass-like progenitor teosinte (*Zea mays* ssp. *parviglumis* Itlis and Doebley) about 7000 years ago in Central America. During domestication – which was brought about by sustained and methodical human effort – an upright, unbranched crop plant with a large, densely packed ear of kernels emerged. Crucially, the kernels had lost the dehiscence zone at their base, which had allowed their dispersal and therefore survival in the wild. However, in domesticated maize kernels, their firmer attachment to the ear made it possible to harvest seed on a large scale and store it for long periods, and this was undoubtedly a crucial factor in the subsequent emergence of the Mayan, Inca and Aztec civilizations. The Native Americans became so dependent on the new crop that they coined the name maize, which means 'that which sustains life'.

Only during the past few years has science come to explain the nature of these events that changed the history of humanity. Because the morphological transition from teosinte to maize represents one of the most astounding and profoundly influential examples of plant breeding ever, it is worth pausing to reflect on what we currently understand of how it was achieved. We now know the identities of some of the critical genes that transformed weedy, unproductive teosinte into high-yielding maize.

6.1.1.1 Five small steps for breeding – one giant step for mankind

The combination of modern quantitative genetic analyses with powerful new genomics technologies has led to remarkable insights into how maize was domesticated. Classical breeders had long known that maize and teosinte were part of the same species complex, and even that the essential differences between maize and teosinte (Figure 6.1) could be explained by a handful of independently segregating genetic factors (Mangelsdorf & Reeves 1938). However, it was not until the early 1990s that quantitative trait locus (QTL) mapping using restriction fragment length polymorphism (RFLP) markers was used to quantify the effects of this small number of 'domestication loci' and to localize them in the maize genome (Doebley *et al.* 1990b). One of these QTL, a locus called *tesointe glume architecture 1* (*tga1*), was genetically fine mapped and found to correspond to a single Mendelian locus, providing the first concrete evidence that selection of allelic forms of single genes had driven the domestication of maize (Dorweiler *et al.* 1993). A second QTL, *teosinte branched 1* (*tb1*), was cloned by transposon tagging (Doebley *et al.* 1997). Interestingly, an alteration in the transcriptional regulation of the *tb1* gene seems to have been heavily selected for during maize domestication, resulting in an extremely low level of nucleotide diversity in its promoter region (Wang *et al.* 1999).

So, it has been shown in principle that the essential morphological characteristics that distinguish maize from teosinte could have been selected in a small number of large steps, and at least in the case of *tb1*, the precise location of the allelic differences

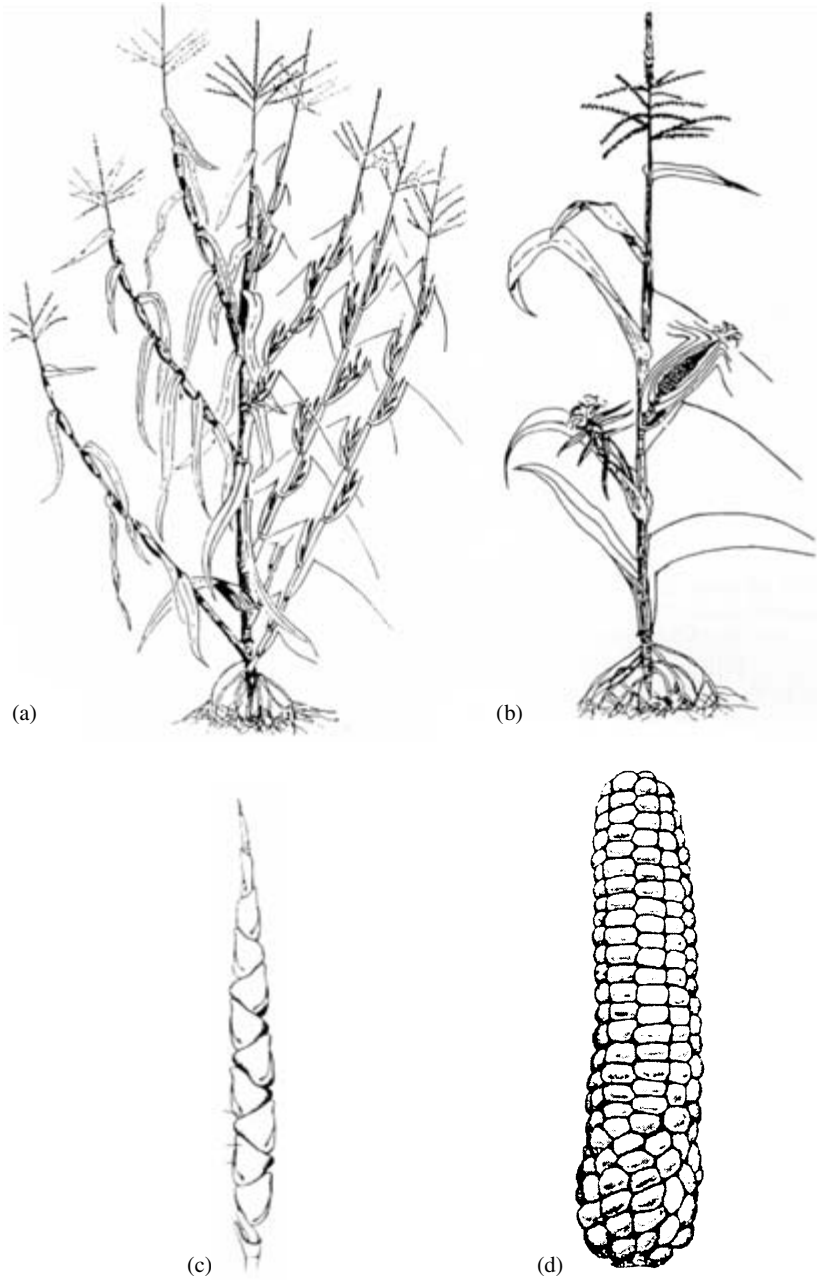


Figure 6.1 Teosinte versus maize morphology. (a) Teosinte plant with long lateral branches bearing both tassels and ears; (b) maize plant with tassels confined to the main stem and the ear on a short lateral branch; (c) teosinte ear composed of fruitcases with hardened outer glumes arranged in two ranks; (d) maize ear with many ranks of uncovered kernels. (After Doebley *et al.* 1990.)

that were actually selected have been demonstrated. But did the domestication of maize really happen in this way or by a more gradual process of evolution?

Again, molecular tools have stepped in to provide new evidence: Matsuoka *et al.* (2002) used single sequence repeat (SSR) markers distributed throughout the maize genome to produce a phylogeny which encompassed both American maize landrace diversity (*Zea mays* ssp. *mays*) and overall diversity within *Z. mays* including subspecies *mexicana*, *parviglumis* and *huehuetenangensis* (the wild teosintes). Since ssp. *mays* forms a monophyletic group within the phylogeny, the results point strongly towards a single domestication of maize. Furthermore, application of a maize mutation model to the data allows estimation of an upper limit for the domestication date of 7186 BC. The overlapping genotypes of certain maize landraces with certain teosinte populations even permit speculation as to the location within the Mexican highlands where domestication was most likely to have occurred.

Slowly but surely, the 'molecular magic' that was practised long before our time is being revealed.

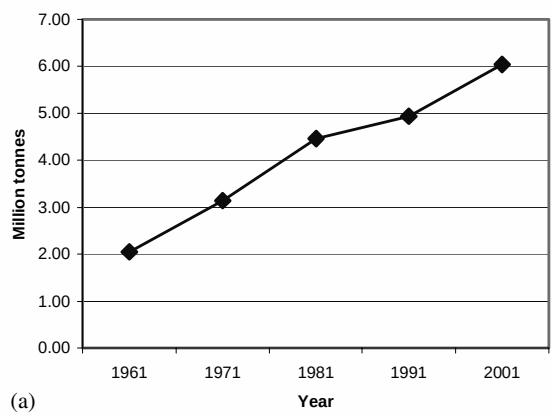
From its centre of domestication in Central America (probably Mexico), maize was spread across much of South America by pre-Inca peoples and was subsequently transported in several episodes from the sixth to the nineteenth centuries by Native American Indian peoples over much of North America. During the spread of maize, its adaptation to local climates and end uses carried on under continuous selection by the peoples who depended on it. Specific mutations in single genes that conferred undreamed of properties on the variant seeds were picked out by observant farmers and multiplied. Hard flinty kernels that were easier to mill into easily stored flours, or soft, sugary kernels that required only light grilling in order to provide a tasty meal – the list is endless. Soon after the 'discovery' of America by Columbus in 1493, maize was brought back to Europe and on to Asia and Africa via the trade routes of the time and has become widely distributed throughout the world, both as a popular subsistence crop and a highly productive cash crop for large-scale cultivation.

Maize currently accounts for 29.3% of world cereal production, having in the last decade overtaken both rice and wheat in terms of tons of primary production (FAO 2002) (Figure 6.2a).

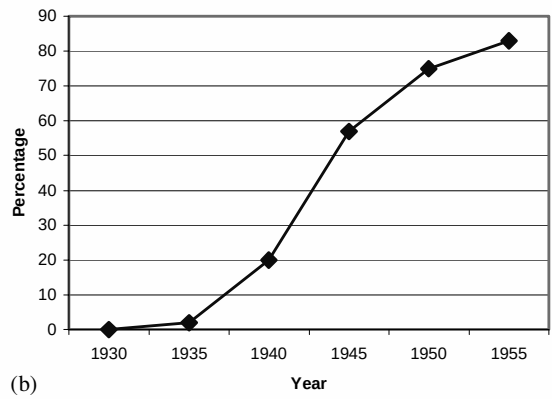
6.1.2 Food and industrial uses of maize

6.1.2.1 Popular maize-based foods

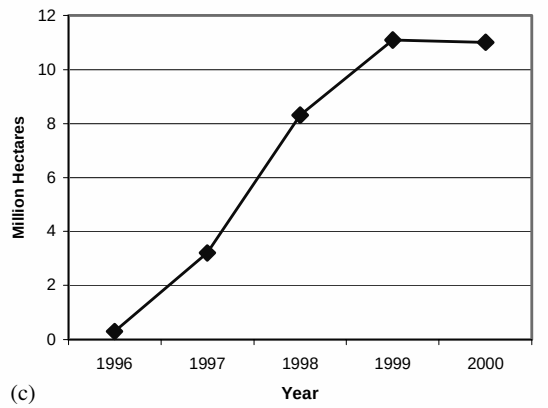
Given its worldwide distribution and scale of production, it is not surprising that maize-based foods have gained prominence in many cultures across the world. In Western cultures, the most familiar maize products consumed are popcorn, cornflakes and sweet corn (which is often grilled or boiled). However, food uses account for less than 2% of maize grown in developed countries, the rest being grown for fodder and industrial uses. In contrast, maize in developing countries (46% of the total world production in 1999) is more usually grown by subsistence farmers for personal consumption. The starting point for most maize-based food preparations is



(a)



(b)



(c)

Figure 6.2 Rising trends in maize production. (a) World production of maize. (b) Percentage of maize area planted to hybrids in the USA between 1930 and 1955 (redrawn from Duvick 2001). (c) Global area planted to GM maize between 1996 and 2000 (data from James 2001).

the removal of the seed coat to expose the starchy endosperm which can be ground wet to produce a dough for immediate use or dried down and milled to produce a flour. This is achieved by heating the whole-grain maize in an excess of 1% lime solution for approximately 1 hour and leaving to stand overnight. The next day, the maize – now referred to as nixtamal – is washed several times with water to remove the loosened seed coats and other impurities, and then recovered. This treatment, which is known as 'lime-cooking', is particular to Mexico and Central America, although today the technology has been scaled up for industrial exploitation and exported to other countries such as the United States (Bressani 1990).

A dough prepared from lime-cooked maize is the main ingredient for many popular dishes such as 'tortillas', 'tacos', 'enchiladas' and 'pupusas' – all of which utilize the baked dough as a base or wrap for a variety of fillings. The dough may also be incorporated into 'atole' (a beverage) and 'tamalitos', which are made by wrapping the dough in maize husks and steam-cooking it until the starch is gelatinized. To make tamales, chicken or pork meat may be added to the gelatinized dough. When the dough is fried and flavoured, it yields foods such as chips and 'chilaquiles'. If the dough is allowed to ferment for 2 days, while wrapped in banana or plantain leaves, it provides a food named 'pozol' from which a number of drinks can be made. Colombian 'arepa' can be described as roasted maize bread without yeast, round in shape, prepared from degermed maize. In Latin America there are many maize-based foods besides tortillas and arepas. Some of these are drinks such as 'colados', 'pinol' and 'macho', which basically are suspensions of cooked maize flour. 'Humitas', which is consumed in Bolivia and Chile, is made from immature common or *opaque2* maize to which is added a number of other ingredients. Humitas is produced from precooked maize flour which resembles the lime-treated 'masa'. From immature maize a sweet, tasty atole of high nutritive value is made. In Egypt, a maize flat bread, 'aish merahra', is widely produced using maize flour to make a soft dough spiced with ground fenugreek seeds. A similar product called 'markouk' is eaten in Lebanon. In Africa, porridges with a moisture content of about 90% are often made. An example is 'ogi', which is eaten all over Nigeria and in parts of Ghana. Maize is also widely used to make beer. In Benin, for example, malt is obtained by germinating the grain for about 5 days; the malt is then exposed to the sun to stop germination. The grains are lightly crushed in a mortar or on a grinding stone, after which the malt is cooked and the extract is strained off, cooled and allowed to stand. After 3 days of fermentation it is ready to be drunk as beer. The critical role of maize in human nutrition has been reviewed in a report published by the Food and Agricultural Organisation of the United Nations (FAO 1992).

It is important to note however, that the majority of maize grown in the world (about 60% in developing countries) is consumed indirectly through its use as animal feed, either as silage or incorporated into processed feeds.

6.1.2.2 Industrial processing of maize

The three most important processed maize products are sweeteners, ethanol and starch – all of which are derived from the starchy endosperm of the maize kernel.

However, oil from the germ (embryo) as well as the gluten protein and fibre that remain after starch extraction are all recovered and used in various by-products. The corn wet milling process separates the four basic components of the maize kernel: starch, germ, fibre and protein. First the incoming shelled maize, which has been stripped from the cob during harvesting, is inspected and cleaned. Then, it is steeped for 30 to 40 hours to begin breaking the starch and protein bonds. The next step in the process involves a coarse grind to separate the germ from the rest of the kernel. The remaining slurry consisting of fibre, starch and protein is finely ground and screened to separate the fibre from the starch and protein. The starch is separated from the remaining slurry in hydrocyclones.

6.1.2.3 Starch uses

Starch obtained from the wet milling process outlined above can be converted to syrup, or it can be made into several other products through a fermentation process. Starch, suspended in water, is liquefied in the presence of acid and/or enzymes which convert the starch to a low-dextrose solution. Further enzymatic treatment continues the conversion process. Throughout the process, acid or enzyme reactions can be halted at key points to produce the right mixture of sugars such as dextrose and maltose for syrups to meet different needs. In some syrups, the conversion of starch to sugars is halted at an early stage to produce low-to-medium sweetness syrups. In others, the conversion is allowed to proceed until the syrup is nearly all dextrose. The syrup is refined in filters, centrifuges and ion-exchange columns, and excess water is evaporated. Syrups are sold directly, crystallized into pure dextrose, or processed further to create high-fructose corn syrup.

Dextrose is one of the most fermentable of all of the sugars. Following conversion of starch to dextrose, the dextrose may be piped to fermentation facilities where the dextrose is converted to alcohol by traditional yeast fermentation or to amino acids and other bioproducts through either yeast or bacterial fermentation. After fermentation, the resulting broth is either distilled to recover the alcohol or concentrated through membrane separation to produce other bioproducts. Carbon dioxide from fermentation is recaptured for sale, and nutrients remaining after fermentation are used as components of animal feed ingredients. Basic consumer necessities such as paper and textiles are major uses for maize starch in sizing, surface coating and adhesive applications. Maize starches and related dextrins (a roasted starch), are used in hundreds of adhesive applications. Special types of starches are used in the search for oil as part of the 'drilling mud' which cools down super-heated oil drilling bits. Other key uses of starch in industry are as flocculating agents, anti-caking agents, dusting power and thickening agents. Literally thousands of supermarket staples are produced using both regular and specially modified starches. Many of today's instant and ready-to-eat foods are produced using starches which enable them to maintain the proper textural characteristics during freezing, thawing and heating. Starches are converted to sugar by breweries and used as the feedstock in beer production. Other starches are the backbone of instant pie and pudding fillings which require little or no cooking compared with traditional formulations.

6.1.2.4 Oil uses

Oil is obtained from the germs (embryos) which are separated from the denser components of the wet-milled slurry in cyclones. The separated germs are then washed to remove any attached starch granules. A combination of both mechanical and solvent extraction processes is used to extract the oil from the germ. About 50% of the US-produced corn oil goes into cooking or salad oil, while corn oil margarines account for approximately 25%.

6.1.2.5 Protein and fibre uses

Many of the by-products of maize processing also return to the food chain through their incorporation into processed animal feed. Through different combinations of steepwater, maize germ residues, fibre and maize gluten, four major feed products can be produced: gluten meal; gluten feed; maize germ meal; and condensed fermented maize extractives (steepwater). Maize gluten meal supplies vitamins, minerals and energy in poultry feeds. Steepwater is a liquid protein supplement for cattle and is also used as a binder in feed pellets, and maize gluten feed provides protein and fibre for beef cattle.

6.2 Twentieth-century maize breeding

As we have seen in the previous section, maize production and consumption patterns are markedly different in developing versus developed countries. Several generalizations can be made regarding this point. Most maize in developing countries is grown in tropical or tropical highland climates in low-input systems by subsistence farmers as a staple food crop. In contrast, only a small fraction of the area used for maize production in developed countries is climatically tropical. It is generally grown intensively on a large scale, and a large part of it is grown as animal fodder. There are therefore equally marked differences in the nature of the breeding programmes that underpin these two modes of production. Since large-scale producers in developed countries systematically buy improved hybrid seed, most of the world's private sector research and breeding effort is directed at their needs and thus to traits most relevant to high-input temperate production systems. This influence also partly extends to public sector research in developed countries. Therefore, when one considers world output in maize research, there is a clear imbalance in the investment put into traits such as herbicide resistance and silage digestibility as opposed to yield stability under marginal conditions or nutritional value of the grain (both of which are issues of huge importance to poorer maize producers who have access neither to expensive agrochemicals nor the variety of other foods needed to create a balanced diet). This distortion in the direction of maize breeding is partly offset by the excellent work of institutions that form part of the UN/World Bank-sponsored Consultative Group on International Agricultural Research (CGIAR) who cater for the needs of the many millions of maize producers who cannot afford to buy improved seed,

thereby reinforcing the efforts made possible by the limited resources available for public research institutions in developing countries.

6.2.1 Germplasm resources

Before considering the outputs of breeding research, let us consider in more detail the basic tools essential to the breeding effort. The first of these is the genetic base material, which will be the starting point for all improvement strategies, whether classical or biotechnology-driven. These can be found from a number of sources.

6.2.1.1 Gene banks

Total diversity among domesticated maize is extremely broad, and includes landraces that have diverged and undergone strict selection on three continents for an endless variety of end uses. In particular, the richness of the maize culture in Mexico and Central America has meant that even today, there are enormously diverse populations of maize under cultivation in the form of landraces. The world's major collections of Mexican maize germplasm are maintained by the International Center for Maize and Wheat Improvement/Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), the Mexican National Institute for Agricultural, Forestry and Livestock Research (INIFAP), and the United States Department of Agriculture's North Central Regional Plant Introduction Station (Ames, Iowa).

The Germplasm Resources Information Network (GRIN) disseminates comprehensive information on US collections of maize germplasm. As of March 2002, GRIN recorded 21 330 accessions from five *Zea* species – *Zea diploperennis*, *Z. luxurians*, *Z. mays*, *Z. perennis* and *Z. nicuraguensis.* The vast majority of these (21 083) represented cultivated maize (*Z. mays* ssp. *mays*). These were principally held in the North Central Regional Plant Introduction Station at Ames, Iowa (18 038 accessions) and the Maize Genetic Stock Centre, Urbana, Illinois (3146 accessions). Other germplasm sources include world collections at CIMMYT, where some 17 000 maize and teosinte accessions are maintained in long-term storage, tropical materials stored at the International Institute for Tropical Agriculture (IITA) and smaller collections at national programmes of different countries around the world. It is also possible to obtain specific germplasm from individual breeders working on maize at public institutions and universities.

It has been estimated that all of maize production is based on about 5% of total genetic diversity. Most maize breeding is in fact, still carried out using familiar sources of tried and tested material, even though the scope for breeding significantly higher yields and novel traits from such material appears to be quite limited.

This situation has been formally recognized in the US, where a coalition formed from the United States Department of Agriculture's Agricultural Research Service, land-grant universities, and industry have come together with the objective of 'widening the germplasm base of commercial hybrid corn in the United States through the introduction and incorporation of novel and useful germplasm gathered from around the globe' in the Germplasm Enhancement of Maize (GEM) project.

Two major concerns of US maize producers have driven this development. First, the diversification in end uses of the maize grain have forced a rethink into where the required grain quality characteristics can be most readily sourced. Since much of the exotic germplasm has undergone selection for many indigenous uses (foods, beverages, etc.) by various cultures, it seems likely that new grain quality characteristics will be found in exotic rather than the narrow-based germplasm now used.

The second major concern with a narrow genetic base is the potential for widespread disease or insect damage due to new diseases or insect species spreading into US corn-growing areas. It is more likely that resistance to these dangers would be found in genetically diverse exotic germplasm sources than in existing breeding material.

6.2.1.2 In-situ conservation

The protection of biodiversity for future generations is certainly catered for in part by the cold storage of seed in germplasm collections across the globe (so-called ex-situ conservation). However, increasingly, research into the feasibility and long-term advantages of in-situ conservation as a complementary and alternative strategy are being examined. Briefly, the advantage of subsidising subsistence farmers to maintain landraces gives the benefit of ongoing selection against current pests, perpetual multiplication of seed and a cultural 'database' of agronomic field data, which accompanies the landrace. In one of the first examples involving maize, CIMMYT is currently conducting a study into the feasibility of in-situ conservation via 'farmer participatory germplasm improvement' – the Oaxaca project (Bellon 2001).

6.2.2 Breeding methods

6.2.2.1 Inbred lines and hybrid maize

The evolution of modern day *Zea mays* ssp. *mays* has resulted in a determinate annual crop which, unlike most other grasses, produces separate (imperfect) sexual flower structures. The female pistillate inflorescences (the ear) are borne at the apex of highly condensed lateral branches emanating from the leaf axils and the male staminate inflorescences (the tassel) terminates the main culm. This physical separation of the sexual organs of maize means that unlike many grass species, most maize kernels are produced by open pollination via wind-borne pollen from neighbouring plants (within a 50-metre radius). Given this situation, it should come as no surprise that until the 1940s the majority of maize seed produced by the commercial seed companies was derived from open-pollinated material (Duvick 2001). Many of the open-pollinated varieties used had their origins in the nineteenth century, and as such had suitably interesting local names such as Lancaster Sure Crop and Reid Yellow Dent. These open-pollinated varieties were capable of yielding up to 3 tons per hectare. Interestingly, Charles Darwin was among several workers who used maize to show that plants derived from open-pollinated kernels were up to 25% taller and more resistant to adverse weather conditions than those derived from self-pollinated kernels (Desmond & Moore 1991). In 1908, G.H. Schull published 'The Composi-

tion of a Field of Maize', and in 1909 this was followed by his 'A Pure-line Method of Corn Breeding.' In these two papers, Schull proposed that the reduction in vigour seen in inbred maize was due to the production of homozygote biotypes (Schull 1908, 1909). Schull also proposed that it should be possible to generate maize lines with superior agronomic traits by crossing suitable inbred-derived material. Occasionally, the results of such crosses were spectacular with some hybrids out-yielding the original varieties by as much as 30%. It is amazing to think that today, almost 100 years on, we still have relatively little idea as to the bases for the phenomenon often referred to as hybrid vigour or heterosis. Heterosis is defined as the difference in yield between a single cross hybrid and the mean of the two inbred parents (East 1936; Meghi *et al.* 1984). Early on in the generation of hybrid maize, it was noticed that certain classes of inbred were more successful in producing high-yielding hybrids when combined with other different specific classes. This simple observation gave raise to the idea of heterotic groups (Moll *et al.* 1965). While heterotic groups are a breeding classification developed by the breeders by trial and error, the classification does have a basis in genetics. Numerous studies using molecular markers have shown that the different heterotic groups tend to associate together (Ajmone-Marsan *et al.* 1998). It is assumed (although not yet proven beyond doubt) that combining two inbreds from two heterotic groups leads to the generation of hybrids with a higher frequency of heterozygote loci than hybrids produced by combining inbreds derived from a single heterotic group. It has been suggested that the generation of hybrid lines with a high degree of heterozygosity gives rise to one or more of the following effects; the addition of alleles with partial to complete dominance, over dominance and/or epistasis (Gardner 1978). Whatever the cause of heterosis, the strategy of inbred/hybrid production is time-consuming. For instance, classical inbreds are generated by self-pollinating individual plants for between five and ten generations. Only then can the inbreds ability to generate high-yielding hybrids be assessed via crosses to other inbred lines derived from different heterotic groups. A significant drawback to this protocol, first suggested by Schull, was that the yield of the early inbred lines was often so low (often less than 50% of the open-pollinated lines they were derived from) that it was not commercially feasible to produce hybrid seed. However, in 1918, Jones suggested that double crosses might be a more feasible method for the generation of hybrid lines (Jones 1918). Whereas single-cross hybrids are derived from a direct cross between two inbred lines, a double-cross hybrid is derived from two pairwise crosses involving a total of four inbred lines. Although more complicated than a single cross, the double-cross-derived hybrids were cheaper to produce yet still out-yielded the original open-pollinated varieties. Based upon Jones's ideas, demand for hybrid maize grew steadily during the 1920s, 1930s and 1940s (Figure 6.2b). The commercial feasibility of hybrid maize became a reality in the 1950s with the re-introduction of single-cross hybrids. In part, this development was a direct consequence of the widespread adoption of the recurrent selection methodology for corn breeding (Hull 1945). Although there are several different forms of recurrent selection as applied to maize, in its simplest form it consists of selecting new inbreds from crosses of the best inbreds derived from previous generations. Such a system

led to the continuous development (selection) of improved inbred lines. Surprisingly, and in part due to recurrent selection, since the 1940s the yield of the individual inbred lines has increased at approximately the same rate as the derived hybrids.

It should be remembered that not all of the yield increase seen during the development of hybrid maize is due directly to hybrid maize. During the period in question there were also consistent developments in agronomic practice – for example, in the use of pesticides, herbicides and the application of nitrogenous fertilizer. The development of new hybrid lines should therefore not be seen in isolation. On the contrary, it is the flexibility of the hybrid maize breeding programmes that allows new lines to be produced that make full use of any developments in agricultural practices. One might ask the question 'what role has molecular breeding played in the process of inbred and hybrid production?' Well to date, not very much! Until the mid-1980s, all inbred and hybrid production was carried out using only the time-honed skill of the breeder. However, as we shall see in Section 6.2.2.3, since the 1980s the production of both inbred and hybrid lines has been revolutionized by genomics-based technologies.

6.2.2.2 Cytoplasmic male sterility

In 1934, only 0.4% of US corn was derived from hybrid corn; by 1945, this had increased to approximately 60%, while by 1956 the figure had increased to over 90% (Figure 6.2b). Coincidentally, the increased use of hybrid maize by the farming community led to the growth of numerous companies entirely devoted to the production of hybrid seed. The simple pressure of market forces on these companies meant that any development in the generation of hybrid seed by more efficient means and at lower costs, would bring the individual company distinct advantages. Since its initial development, the production of hybrid seed corn had required the detasselling of the female inbred parent. In the US, this event was considered a valuable source of student vacation money, in that detasselling was best undertaken manually a few weeks before pollen shed. Thankfully for the students, this period coincided with the student summer vacation! For many years before the 1970s, cytoplasmic male sterile (CMS) lines had been used on a small scale for the production of hybrid varieties (Duvick & Noble 1978). These lines had the advantage that they did not require detasselling and so were cost-effective. The first case of CMS in maize was described by Rhoades in 1931. CMS appears to have occurred many times in various maize lines; however, it appears that many of these events represent the same underlying mechanisms. From all the various types of CMS there appear to be three different forms termed T (Texas), S (USDA) and C (Charrua). If CMS lines have the potential to reduce costs during hybrid production why are they not in much more widespread use? The answer can be found in the disastrous consequences of the widespread use of T-CMS during the late 1960s and early 1970s. The events of 1970 are well documented (Duvick & Noble 1978). Briefly, in 1970 approximately 80% of the US maize crop was based on T-CMS. Early in the 1970 growing season a new race (now called race T) of *Helminthosporium maydis* (Southern corn leaf blight) was noticed which infected the southern corn belt corn fields. During the

summer of 1970 the disease moved north, infecting the crop and reducing yield by approximately 15%. Further research on the new race showed that it was especially virulent on T-CMS lines. Remarkably, the seed companies reacted very quickly to the disaster and by the following year virtually all the seed offered for sale were in normal cytoplasm.

With the advent of genomics-based technology however, it appears that male sterility is set for something of a revival (see Section 6.2.4).

6.2.2.3 Molecular markers and marker-assisted selection

Maize was one of the first plant species to be analysed using molecular markers. For instance, Sheridan's *Maize for Biological Research* (1982) reported a list of over 20 isoenzymes, which were in regular use as molecular markers. However, highly polymorphic isoenzymes are relatively rare and they were soon replaced by RFLPs during the early 1980s. Although RFLPs have now been superseded by SSRs (Sharopova *et al.* 2002) and single nucleotide polymorphisms (SNPs, but pronounced SNiPs: Brookes 1999; Batley *et al.* 2002), the importance of RFLPs, and their influence on molecular breeding of maize should not be under-estimated (Tsaftaris 1996). A concerted effort by the academics and private sector and the adoption of common mapping populations rapidly led to the development of a dense RFLP genetic map (Davis *et al.* 1999).

In maize, RFLPs were shown to be highly polymorphic, possibly due to the high density of transposons and retrotransposons (SanMiguel *et al.* 1996). In the 1980s, these factors encouraged all of the larger breeding companies to invest heavily in RFLP technology, such that it was common to have teams of up to 20 technicians solely dedicated to carrying out the RFLP procedure. Given this input, via RFLPs it became possible for companies to examine several tens of thousands of lines in a single season. Interestingly, the use that RFLPs were put to diversified in the different breeding companies. For instance, the more successful breeding companies were slow to introduce RFLPs into their direct breeding programmes, for instance in the analysis and selection of QTL. In hindsight, it is difficult to understand this logic, but given the success of their existing breeding programmes (driven entirely by conventional breeding technologies), instead these companies focused their RFLP effort onto variety identification, the protection of intellectual property and the examination of possible essentially derived varieties (Smith 1998). The ability to fingerprint proprietary maize inbred and hybrid lines is of considerable commercial interest to the large seed companies. For instance, in a now famous case, Pioneer Hi-Bred was awarded $46 million by the US supreme court against Holden Foundation Seeds Inc. for the alleged illegal acquisition of inbred lines belonging to Pioneer. A considerable amount of the evidence against Holdens was provided by DNA fingerprinting, a development which provided a timely reminder of the dangers of seed companies acquiring seed by dubious means.

6.2.2.4 Candidate genes: the next generation of molecular markers

What of the future of molecular markers for maize breeding? The answer can be found by remembering what molecular markers are; quite simply they are (chromosomal) markers for genes which, when in the correct allelic configuration, generate a plant with favourable agronomic traits. However, what if the markers are the genes themselves? Ever since the late 1990s private companies have, via either in-house efforts or via collaboration with specialized sequencing companies, been sequencing maize genes in the form of expressed sequence tags (ESTs). The public maize EST programmes began in the early 1990s as a relatively minor sequencing project (Keith *et al*. 1993). However, these efforts now seem to show a significant degree of foresight. In 1998, via a significant grant from the National Science Foundation (NSF), the public sector finally joined the maize EST race with its own effort. As of May 2002, the public database contained 153 214 maize ESTs (zmDB: http://www.zmdb.iastate.edu/zmdb/nsf_grant.html). Many of these sequences are derived from the maize gene discovery programme (Gai *et al*. 2000), which was funded in the 1998 round of NSF awards. Although it is difficult to gauge the number of maize ESTs in private company databases, it is probable that all of the big three companies (Monsanto, Syngenta and Pioneer-Dupont) have between 500 000 and 1 000 000 maize ESTs in their respective databases. One important aspect of the public NSF-funded programme is that many different (more than nine) cDNA libraries were included in the analysis. Hence, the current database – while not being a complete reflection of the maize transcriptome – is a useful starting point for the discovery of genes with specific expression profiles. Recent analysis of the 153 214 individual ESTs suggests that they form 31 408 unique contiguous sequences (contigs). These contigs contain 17 384 clusters and 14 024 singletons – a figure which suggests that a significant number of maize genes are expressed at a relatively low rate. Therefore, whilst the current 31 408 contig figure is short of the estimated 50 000 genes in the maize genome, it does represent a useful starting point in the search for candidate genes underlying agronomically important QTLs. Candidate genes are, as their name implies, genes which both map to the region known to contain an inherited trait and which based on sequence similarity could encode a protein which might provide the required function. In many cases the true candidate is easy to spot. For instance, it is not too difficult to imagine how genes involved in lignin biosynthesis, for instance *cad* or *pal*, might be involved in stalk strength.

Once candidate alleles or groups of alleles have been shown to be associated with specific agronomic traits, the information provides a convenient means to screen for the desired allelic combination via sequence-based genotyping. Sequence polymorphism between different candidate gene alleles can take many forms; for example, it can be due to the insertion or deletion of multiple bases within either the untranslated or translated regions or it can be due to SNPs. Insertions, deletions (indels) and SNPs can be used for sequence-based genotyping of maize. Mogg *et al*. (2002) have shown that maize ESTs contain a significant number of SNPs and indels; for instance, for the average EST, SNPs could be found every 40 base pairs (bp). Given that the maize genome is estimated to be 2.5×10^9 bp in size, this means that there is the potential for

up to 62 million SNPs when two or more inbred lines are compared. This relatively high density of SNPs in the maize genome brings about the possibility of identifying candidate genes via linkage disequilibrium (Brookes 1999). The dense genetic maps (up to one marker per 40 bp) possible with maize SNPs allow genome scans for linkage disequilibrium of the SNPs to be studied in association with complex phenotypes. In 1998, the NSF awarded $2.2 million to a consortium of American University laboratories to develop SNPs for genes involved in quality traits in maize (http://www.nsf.gov/bio/pubs/awards/genome98.htm). In 1999, the NSF awarded a further $2.9 million to Iowa State University to develop complementary SNPs-related technology (http://www.nsf.gov/bio/pubs/awards/genome99.htm).

6.2.2.5 Foreground and background selection

One of the most straightforward molecular marker applications carried out today by the molecular breeder is that of background and foreground selection. Often, the molecular breeder is required to introgress a single gene, or small number of genes (say herbicide resistance and/or single gene disease resistance) into an elite genetic background. This is usually carried out by first crossing the two inbred parental lines, followed by continuous backcrossing to the recurrent inbred parent. At each stage, the breeder is required to both select for the trait being introgressed and as high a percentage as possible of the elite background. Before the introduction of molecular markers, it was often necessary carry out six or seven backcrosses to the recurrent parent in order to guarantee a sufficient and consistent amount of elite genetic background. However, with the introduction of molecular markers and their ability to identify individual plants containing a larger percentage of the recurrent parent than the average value, the number of backcrossess can be reduced to perhaps three or four. This simple development means that new lines can be created with specific traits in two rather than four years (Eathington *et al.* 1997). The availability of large numbers of markers – and perhaps even the trait genes themselves – means that the physical distance between the marker and the trait gene is kept to a minimum and linkage drag (i.e. the associated genes, or more correctly alleles, which might confer deleterious traits) are reduced to an absolute minimum. Utilization of at most two polymorphic markers surrounding the trait gene is all that is required to select for the trait being introgressed. However, it should be noted that multi-locus marker technology, such as amplified fragment length polymorphisms (AFLPs) or multiplexed microsatellites are required to select for the recurrent background elite parent (Pejic *et al.* 1998).

6.2.3 Breeding targets

If you speak directly to a maize breeder, it is probable that they would suggest that the only truly profitable target for breeding is yield. In the case of maize, yield refers to the total amount of grain produced from a given area. However, with the emergence of 'speciality maize' – that is, maize bred for a specific purpose (for instance, low lignin or high amylose) rather than for the quantity of grain produced – breeders now

have to rethink their strategy. It is important to realize that this process is still in its infancy; during the coming decades an increasing amount of maize will be bred for specific end users and for highly specialized purposes.

In the past, breeding for increased yield has been focused on maximizing hybrid vigour through the study of the combining ability of elite inbreds from different heterotic groups as discussed in the previous section. We will therefore restrict the discussion of breeding targets to a consideration of some examples where strategies for rational improvement through a better understanding of the underlying molecular biology has become possible.

6.2.3.1 Example 1: Starch synthesis

The component with the greatest concentration in the maize kernel is starch (~80% dry weight). The *Waxy* (*Wx*) gene in waxy maize has been shown to control the amount of amylopectin starch in the endosperm (up to 100%) with very low amounts of amylose (Creech 1965). This gene has now been cloned (Shure *et al.* 1983). Other genes cause an increase in reducing sugars and sucrose, for example, the *Sugary* (*Su*) gene produces relatively high amounts of water-soluble polysaccharides and amylose. Maize kernels containing this gene are sweet and are important for canning. Their starch content and quality also have nutritional implications, since some starch granules have low digestibility while others have high digestibility, as demonstrated by Sandstead *et al.* (1968). This gene, too has been cloned (James *et al.* 1995) and the molecular basis for variation at the *sugary* locus has recently been elucidated (Dinges *et al.* 2001). However, perhaps because allelic variation at these loci has been long used (with great success) by the classical breeders, knowledge of the molecular basis for these important traits has not (until present at least) been exploited for engineering maize with modified starchy endosperm.

6.2.3.2 Example 2: Quality protein maize (QPM)

The low protein quality of maize stems mainly from the deficiency in the protein of the essential amino acids lysine and tryptophan. Mertz *et al.* (1964) found that the *opaque2* gene significantly increased the lysine and tryptophan content in maize endosperm. This gene also reduced the leucine level, giving a better leucine-to-isoleucine ratio. This gene was eventually cloned, and shown to encode one of the first plant transcription factors (Hartings *et al.* 1989). Since 1989, the molecular mechanisms underlying the action of *opaque2* have been elucidated in great detail (e.g. Yunes *et al.* 1994; Vicente-Carbajosa *et al.* 1997; Kemper *et al.* 1999). However, despite its obvious potential to increase the nutritional value of maize for human consumption, *opaque2* maize could not until recently be used widely since incorporation of the gene conferred deleterious effects on yield and increased susceptibility to pests.

The situation changed only after decades of painstaking breeding research conducted at the CIMMYT which led to the identification of essential modifier genes that counteracted the negative effects of deploying the *opaque2* gene. This research eventually yielded the so-called quality protein maize (QPM), which performs ag-

ronomically as well or better than most common maize varieties. The World Food Prize was awarded to CIMMYT breeders in 2000 for this achievement.

6.2.3.3 Example 3: Chill tolerance

Chill tolerance is of considerable importance to maize growers, especially in Europe where the cool wet springs pose special problems. The cool wet springs of Northern European countries limit the growth of grain maize, due to the influence that cool conditions can have on final grain yield (Creencia & Bramlage 1971; Frei 2000). Low temperatures (<8°C) during germination and shoot emergence have been shown to have a significant effect on nutrient uptake, cell division and photosynthesis. However, Bertoia *et al.* (2002) have shown that different maize lines have significantly different responses to cool conditions. Anderson *et al.* (1995) suggested that several genes – including catalase and peroxidase – are expressed during chilling, suggesting a role for these genes in the chilling process. Lipid transfer proteins have been implicated (i.e. they can be considered to be candidate genes) in chill tolerance due to their role in determining the level of phospholipids and their influence on membrane lipid composition (Bishop 1986). Molecular markers have recently been used to try to identify regions of the genome response for controlling the plant's ability to grow under cool conditions (Marton 2000). To date however, the majority of research performed in an attempt to improve the responses of maize to cool conditions has been carried out using existing genetic variation.

6.2.4 'First-wave' maize biotechnology

The most radical innovation of twentieth century maize improvement has been the introduction of genetic engineering to transfer genes encoding novel traits directly into maize. A variety of methods have been developed for maize transformation, including *Agrobacterium*-mediated gene transfer (Gould *et al.* 1991), silicon carbide 'whisker' transformation (Frame *et al.* 1994) and biolistic methods (Frame *et al.* 2000). Although these methods have all been in use for some time, it appears that maize transformation is still difficult for all but a few laboratories.

On the basis of early developments in the field of maize transformation, a 'first wave' of genetically modified (GM) maize was brought out incorporating transgenic modifications that either appealed to producers (e.g. herbicide or pesticide tolerance) or made hybrid production more efficient and therefore more profitable (e.g. male sterility). All represented single gene gain-of-function modifications to the maize genome, and for the most part did not employ native maize sequences.

6.2.4.1 Bt maize

Bacillus thuringiensis is a soil bacterium the spores of which contain a crystalline (Cry) protein. In the insect gut, the protein breaks down to release a toxin, known as a delta-endotoxin. This toxin binds to and creates pores in the intestinal lining, resulting in ion imbalance, paralysis of the digestive system and, after a few days, insect death. Different versions of the *Cry* genes, also known as 'Bt genes', have been

identified. They are effective against different orders of insects, or affect the insect gut in slightly different ways.

Bt maize incorporates transgenes that result in the expression of the *cry* protein in pollen and in all vegetative tissues of the plant, rendering them toxic to the larvae of various lepidopteran pests.

Bt insect-resistant maize is used primarily for control of European corn borer, but also corn earworm and Southwestern corn borer and is sometimes marketed under the trademarks YieldGard® and StarLink®.

6.2.4.2 Herbicide-resistant maize

Herbicide-tolerant maize incorporates transgenes providing tolerance to the broad-spectrum herbicides Roundup® (glyphosate) or Liberty® (glufosinate). Either of these chemicals can provide total weed control provided that there has been no opportunity for resistant weeds to be selected. These varieties are commonly marketed as Roundup Ready® or LibertyLink® varieties.

6.2.4.3 Engineered male-sterile maize

In T, S and C-based CMS (see Section 6.2.2.2), both the original CMS mutation and the restorer genes are the natural products of plant breeding. However, in 1995 Plant Genetic Systems announced that it had developed an artificial sterility system based upon the *barnase* (a ribonuclease) and *barstar* genes (SeedLink™). When placed under the control of an anther-specific promoter, the *barnase* gene produces an RNase that degrades cellular RNA and inhibits the production of viable pollen and hence leads to male sterility. Introduction of the *barstar* gene inhibits the action of the barnase gene product and hence restores viable pollen production. Therefore to produce hybrid seed, one parent inbred line, made male-sterile via the introduction of the *barnase* gene, is crossed (in the field this can be done using simple alternate rows) with a second parent inbred line containing the *barstar* gene. Hybrid seed is then only collected from the male-sterile inbred parent. It should be noted that in this system, because the original plants were homozygous for their respective components, then the resulting hybrid seed would be heterozygous for both *barnase* and *barstar.* Therefore, if F_2 seed is collected from the SeedLink lines, 25% of the plants will be male-sterile. This has serious consequences for farmers in developing countries.

6.2.4.4 Production of GM maize

Being of such importance to the world agricultural economy, it is hardly surprising that various maize products featured strongly among the first generation of transgenic plant varieties to be developed. Once approved and on the market, the take-up and extensive growing of transgenic maize has been extremely fast. In 2001, four principal countries grew 99% of the global transgenic crop area. The USA grew 35.7 million hectares (ha) (68% of the global total), followed by Argentina with 11.8 million ha (22%), Canada 3.2 million ha (6%) and China 1.5 million ha. Globally, GM corn at 9.8 million ha (19%) was the second most widely planted GM crop after GM soybean, occupying 33.3 million ha in 2001 (63% of global area). During the

6-year period between 1996 and 2001, herbicide tolerance has consistently been the dominant trait, with insect resistance second. *Bt* maize, occupying 5.9 million ha (equivalent to 11% of global transgenic area and planted in six countries) was the second most dominant GM crop/trait combination in 2001 after herbicide-tolerant soybean. The global trend in adoption of GM maize between 1996 and 2000 is shown in Figure 6.2c.

6.3 Maize genomics in the twenty-first century

As we have seen, maize breeding has become more versatile and efficient thanks to the benefits of new technologies such as marker-assisted selection and genetic transformation. However, there has been no fundamental revolution in the breeders' art despite glimpses of the potential for such a development. Improvements in a huge majority of traits still derives from selection of a phenotype whose molecular basis is not any the better understood for being tagged with an anonymous molecular marker. Furthermore, it is clear that the first direct genetic modifications of the maize genome resulted more from advances in microbial genetics than from the basic understanding of plant molecular biology and therefore represented an optional addition to the breeder's toolkit rather than a replacement for older methods. It can be argued that the reason for this is that while we may understand how individual genes are encoded, we still have little detail of the complex genetic networks that these genes belong to.

However, there have been notable advances in the study of the genomes of higher organisms in recent years, including some critical developments in plant genomics. Indeed, maize geneticists have not been slow in developing genomics tools for maize.

From the point of view of maize breeding and improvement, the most important tasks for genomics are to provide:

- A complete and accurate catalogue of maize messenger RNAs and their expression patterns in different tissues and in response to various environmental challenges (transcriptomics).
- A complete catalogue of proteins found in various tissues, stages of development and physiological states (proteomics).
- A primary function for each expressed maize gene (functional genomics).
- The ultimate tool for locating the gene sequences underlying qualitative and quantitative functional differences between maize varieties – the entire genome sequence of a reference genotype (structural genomics).
- Extensive data on polymorphisms within and between varieties (evolutionary genomics).
- A bioinformatics infrastructure capable of drawing all appropriate relationships between these different categories of data (bioinformatics).

We will review each of these areas in turn (with the exception of proteomics, which is relatively in its infancy as applied to maize) and show how together they will provide a platform for systematic genome-wide discovery of maize gene function.

6.3.1 Maize gene function discovery

6.3.1.1 Transcriptomics

ESTs have been introduced above in the candidate gene class of molecular marker, where it was pointed out that ESTs of unknown function are relatively little use to the maize breeder. They may have some use as molecular markers if they can be shown to map to regions which have been shown to contain genes whose proteins confer favourable agronomic traits. Even in the case where the function of the EST is known, in most cases it will be of little value to the plant breeder. This fact often comes as a surprise to the academic researcher who may have taken several years of dedicated work to uncover a putative function for the EST. However, plant breeders are not so much interested in genes as in alleles! It is the combination of favourable alleles which determines favourable agronomic traits, not the genes. That said, it will be many years, even for maize, before we are in a position to mix and match alleles, let alone genes. Therefore, despite the previous statements, during the next 5 to 10 years the primary task of the molecular geneticists appears to be to determine the function of the 50 000 or so maize genes. Microarray technology is currently the first step in this process. If one can determine where and under what circumstances a gene is expressed, then one is some way towards determining its function. At conferences, this has often been referred to as 'guilt by association'. Fortunately, for maize geneticists high-density EST arrays are routinely available at a reasonable price (~$140 for two slides) from the University of Arizona via the Maize Gene Discovery Programme (http://gremlin3.zool.iastate.edu/zmdb/microarray/). The arrays come in various formats, but they usually consist of approximately 8000 ESTs derived from a variety of sources (Fernandes *et al.* 2002). It can only be a matter of time before their use is commonplace for the determination of gene expression profiles.

6.3.1.2 Functional genomics

As we have seen, current molecular technology has led to the generation of a large number of molecular markers, and many of these markers can be converted into candidate genes. However, currently, for every maize gene whose function is known (or thought to be known) there are ten genes for which we have little or no idea of function. If we had an idea as to the function of these unknowns then, for any given physically mapped trait, the number of candidate genes would decrease significantly. To achieve this next goal, technologies are required that help plant geneticists to predict the function of genes based solely on their sequence. To confirm the function of a specific DNA sequence, transposon tagging has been widely employed by maize geneticists. Knowing when and where a gene is expressed often gives the researcher an idea as to function; however, it is often only by removing the gene product that a putative function can best be described. Again, maize is almost the ideal organism

in which to approach so-called gene knock-out technology. Gene knock-out technology relies on the ability of transposable elements to insert into the gene, thereby removing its ability to generate a functional RNA and/or protein. The resulting plant can then be observed for any obvious or not-so-obvious phenotype. This approach has been referred to as the gene machine approach (Meeley & Briggs 1995). During the past 10 years, maize transposable elements have been used extensively as 'gene machines' (Das & Martienssen 1995; Hanley *et al.* 2000; Edwards *et al.* 2002). Such an approach has several advantages, including:

- the rapid production of mutant plants for functional analysis;
- the generation of molecular tags which permit the rapid isolation of the mutated gene; and
- the possible provision of a reversible phenotype aiding molecular analysis of the trait.

Whilst such approaches as those described here have been used for the past 10 years, and have generated much data on the function of a small number of genes, in so far as the authors are aware no plant material from such work has formed the basis of commercial breeding. In fact, it is probable that commercial maize breeders would be extremely reluctant to use transposon-tagged material in their breeding programmes for fear of contaminating their breeding stock with variation-generating active transposons.

6.3.1.3 Genome sequencing/structural genomics

Unfortunately, even extensive information from transcriptomic and proteomic studies leaves a critical gap in our understanding of how a given gene works. In order to understand how the gene is regulated and to have access to the upstream and downstream DNA elements that mediate complex cellular control mechanisms, raw genomic sequence is required, which tells the story of the genomic context in which the gene occurs (which may be important in determining its overall transcription state through control of chromatin structure) as well as revealing potential regulatory elements such as promoters, enhancers and alternative intron splice sites.

In this section, we will review the recent history of exploration of the maize genome, and will examine the prospects for achieving the long-term goal of obtaining the full genome sequence of maize.

Important advances in our understanding of how higher plant genomes are organized have come from maize. The development of a maize yeast artificial chromosome (YAC) library resource (Edwards *et al.* 1992) permitted maize geneticists to look at the structure of non-coding maize DNA which makes up approximately 70 to 80% of the genome. Notably, a seminal study of the *adh1* (alcohol dehydrogenase 1) region revealed that nested arrays of transposable element – chiefly retrotransposons – made up a large proportion of the intergenic DNA (SanMiguel *et al.* 1996) and were derived from successive waves of invasion (SanMiguel *et al.* 1998). More recent studies of genome structure found evidence for limited stretches of high

gene density among blocks of retrotransposon arrays (Feuillet & Keller 1999; Fu & Dooner 2000; Fu *et al.* 2001). These results helped to explain why the diploid maize genome is so large (2500 Mb) compared with that of other diploid plant species such as *Arabidopsis* (125 Mb) or rice (420 Mb), and suggested that maize gene density can vary significantly from one region to another.

Another hugely significant class of repetitive DNA – miniature inverted-repeat transposable elements (MITEs) – were also discovered in maize (Bureau & Wessler 1992). These constitute a significant proportion of the maize genome and are probably important sources of genic polymorphism because of their widespread occurrence in and near maize genes (Zhang *et al.* 2000). In fact, both of these types of repetitive DNA have been identified in a number of plant genomes and their role as agents of evolutionary change recognized (reviewed in Wessler *et al.* 1995).

Although the discovery that large blocks of maize repetitive DNA were dispersed throughout the genome was bad news for those who hoped to carry out positional cloning, physical mapping or genomic sequencing, increased understanding of the nature and dynamics of different classes of repetitive DNA allowed them to be harnessed as convenient and informative fingerprinting tools (Edwards *et al.* 1996; Casa *et al.* 2000). Furthermore, the new knowledge allowed the development of algorithms which overcame difficulties in assembling fingerprint and sequence overlaps in repetitive regions (Lin *et al.* 2000; Yu *et al.* 2002).

These advances in understanding maize genome structure were accompanied by rapid progress in resource building. Bacterial artificial chromosome (BAC) libraries were created (O'Sullivan *et al.* 2001; Tompkins *et al.* 2002) and large EST and SSR resources accumulated through extensive co-operation among the maize community (see Section 6.3.2). A large NSF-funded project aiming to produce a physical contig map integrated with a high-resolution SSR genetic map is now well underway led by the Coe laboratory in Missouri (Coe *et al.* 2002). As of 25 March 2002, 10× genome equivalents of maize BACs (203 765 clones) from inbred line B73 had been fingerprinted and organized into 6693 contigs and 8478 singleton BACs. When complete, 450 000 BACs will have been fingerprinted, assembled and integrated with the genetic map. Progress on the final product – the integrated map – can be viewed at http://www.maizemap.org/.

Given the success and imminent completion of the integrated map, the community is collectively looking forward to the next step for which funding is currently being sought for a US-led genome sequencing effort (Bennetzen *et al.* 2001). Initially, sequencing of the coding portion of the maize genome is being targeted, based on sequencing of gene-enriched libraries (for example, Rabinowicz *et al.* 1999), and there is at present considerable optimism that the momentum created will eventually lead to a complete maize genome sequence being released into the public domain.

Although the benefits of the long-term goal of whole genome sequencing are easy to see, even an integration of physical and genetic maps would be of enormous interest to maize breeders. This would allow any marker–phenotype associations emerging from QTL or association studies to be immediately related to a specific

physical region of the genome which could be further explored in pursuit of specific genes or even more tightly linked markers.

6.3.1.4 Evolutionary genomics

An essential task of genomics if it is to have any hope of being applied to breeding is to catalogue on a large scale the genic polymorphisms within and between maize varieties. Maize evolutionary genomics – as this activity is sometimes labelled – is emerging as an efficient new tool in assigning function to candidate genes. It uses association testing, already extensively used in *Drosophila* and human genetics, to correlate polymorphisms in candidate genes with phenotypic variation. The method outlined in Figure 6.3 has been used by Buckler and co-workers to associate variation in the *dwarf8* gene with variations in flowering time (Thornsberry *et al.* 2001), and the sweet maize phenotype with a single nucleotide polymorphism in the *sugary1* gene. Using similar methodology, researchers at Dupont reported to the 43rd Maize Genetics Meeting (see Doebley 2001) that the familiar yellow colour of maize kernels is explained by an insertion just upstream of the *yellow1* gene, which may result in the up-regulation of the gene encoded by *yellow1*. These are of course just illustrations of how the method works, and the genes selected for these proof-of-concept experiments were already characterized, underlying well-studied traits. However, when extended to a much larger number of anonymous genes of unknown function, association testing seems to offer a revolutionary new way to determine what newly discovered genes do.

6.3.1.5 Bioinformatics

Rather than attempt to review the burgeoning field of bioinformatics, we will simply point out two major data repositories that have been developed specifically for maize genomic data. The oldest and most familiar public repository of maize genetic data is Maize dB (http://www.agron.missouri.edu), where a relational database of maize maps, mutants, markers, references and even maize researchers has been maintained since 1991. Maize dB has recently spawned a new website (http://www.maizemap.org) devoted to presentation of the new integrated genetic and physical map (iMap). iMap presents the genetic map, associated BAC contigs of the developing physical map, and access to dynamic links for underlying contig and marker data.

A sequence-indexed database of transposon insertion lines (rescueMu) as well as a large public collection of ESTs and associated microarrays can be accessed through ZmdB (Gai *et al.* 2000). Of interest also at ZmdB are a number of useful maize gene prediction and spliced alignment tools (AAT, GeneSeqer, SplicePredictor, MyGV).

6.3.2 Comparative genomics

Pure functional genomics approaches, as described above, provide an obvious route to knowledge of the molecular components of biochemical pathways of interest.

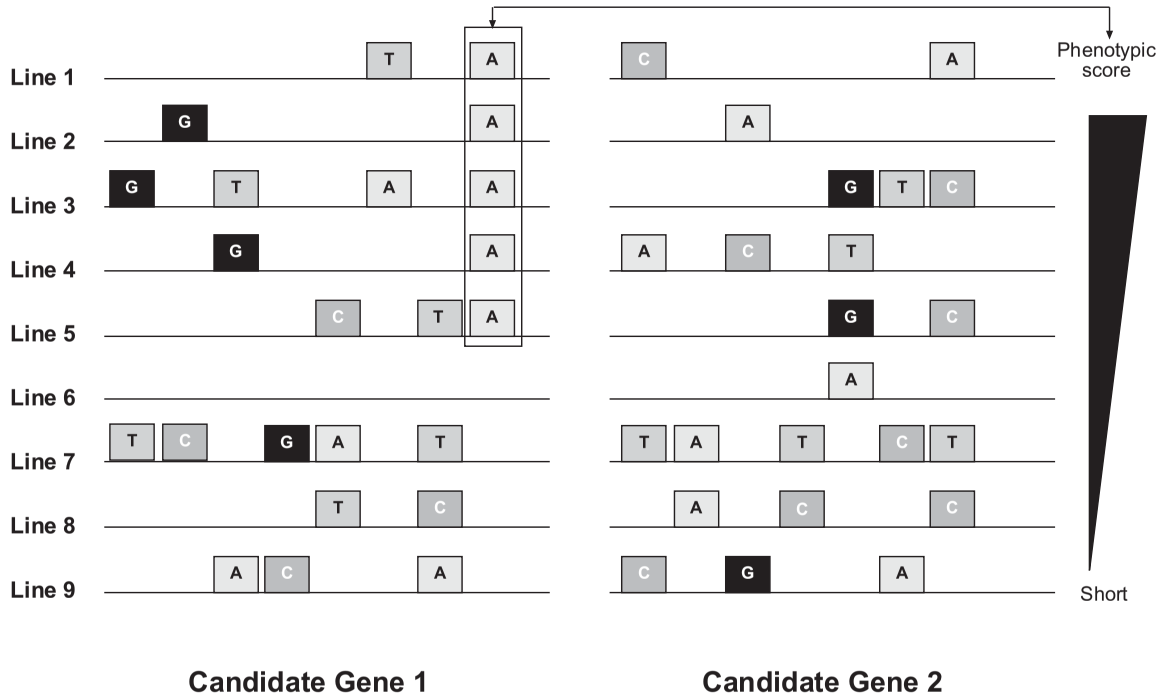


Figure 6.3 Principle of association testing applied to maize. A large number of maize lines representing a suitable cross-section of maize diversity are evaluated for any quantitatively variable phenotype of interest. In this schematic example, nine maize lines are arranged in descending order for their phenotypic score (plant height for example). Genes which are postulated to influence plant height are amplified from each line using PCR, and their nucleotide sequence obtained. Only nucleotide positions that vary from the established consensus are shown, and in this case a clear-cut association between the 'A' in the last variable position of candidate gene 1 and tallness. Advanced statistical methods are used to eliminate false associations, and to detect more subtle trends in the data. Advantages of this approach are that the same set of lines can be used to investigate almost any trait of interest.

However, it will be some time before the functions of even a reasonable fraction of maize genes can be confirmed by transposon tagging or transformation-based methods for modulating their expression. After all, the *Arabidopsis* community have recently set themselves the aim of a complete functional annotation of the *Arabidopsis* genome by 2010 (Chory *et al.* 2000), having released a finished genome sequence in 2000. Draft genome sequence for both *indica* and *japonica* subspecies of rice has been released very recently (Goff *et al.* 2002; Yu *et al.* 2002), so it would not be unreasonable to expect that until maize structural genomics has had time to catch up, gene discovery for maize application will be strongly influenced by comparative approaches.

6.3.2.1 The grass genome

Comparisons between the complete genomes of any two living organisms are valid at some level – even plant–animal comparisons allow powerful conclusions to be drawn about the scope of evolutionary change (Meyerowitz 2002). However, if the objective is to draw inferences about species of interest from related model species by exploiting similarities in genetic programmes, then the closer the two are related, the more valid the comparison.

Cereal geneticists realized the benefits of comparison when the first detailed RFLP map of sorghum was generated using probes exclusively from maize (Hulbert *et al.* 1990). However, once extensive maize–rice synteny (Ahn & Tanksley 1993) was shown and comparisons among the Triticeae had been completed (Devos *et al.* 1993), the door was open for the grass family (despite some 60 million years having passed since a common ancestor) to be declared a single genetic system (Bennetzen & Freeling 1993) with their respective genomes aligned in the now-famous series of concentric circles (Moore *et al.* 1995).

Two grass species are of special relevance to maize improvement. The first is sorghum, which is thought to have diverged from maize approximately 20.5 million years ago (Gaut & Doebley 1997). The sorghum genome at 750 Mb is about 3.5-fold smaller than maize, and yet its gene content and order has long been established to be highly conserved with its larger relative (Hulbert *et al.* 1990).

When *adh1*-orthologous regions of maize and sorghum were compared by cross-hybridization experiments, a number of low-copy number sequences were conserved in their structure and their order, while highly repetitive sequences were divergent (Avramova *et al.* 1996). Comparison of the *sh2/a1* interval between sorghum and rice not only shows that the genes appear in the same order, but also that they are separated by only about 20 kb in sorghum and rice, while they are about 140 kb apart in maize. These observations provide at least two immediate applications to maize improvement strategies: first, the possibility to obtain the genome sequence of a close relative of maize which could reveal up to an estimated 92% of maize genes that share a high degree of similarity with sorghum at less than one-third of the cost of a maize genome sequencing project; and second, the completion of a sorghum EST project will allow determination of the precise degree of overlap between maize and sorghum gene content as well as pinpointing the novel genes unique to the maize

lineage. These goals may not be so far away from realization, as a project which is already underway as part of the US Plant Genome Program (http://www.nsf.gov/bio/pubs/awards/genome01.htm) promises to deliver both an integrated sorghum physical and genetic map as well as a significant sorghum EST database. The integrated sorghum map has been proposed as a foundation upon which systematic cross-referencing of several grass genomes could be based (Draye *et al*. 2001).

By comparison with sorghum, rice is somewhat more distant from maize, having diverged from a common ancestor between 50 and 60 million years ago (Doebley *et al*. 1990a). However, the advanced state of rice genomics, with a draft sequenced genome (Goff *et al*. 2002; Yu *et al*. 2002), comprehensive transcriptome and exemplary breeding resources will mean that for some time to come, the rice–maize comparison will influence research strategies across the globe. It is extremely likely that many advances in understanding of monocot-specific biology will come from studying the rice model and that directly or indirectly, this will have a large impact on maize improvement. For example, Goff *et al*. (2002) described the alignment of maize QTL for grain yield with draft rice genome sequence in a region with demonstrated macro- and micro-synteny. This comparison identified approximately 100 unmapped maize cDNAs, which become candidate genes for the QTL.

6.4 A new paradigm for molecular breeding of maize

Armed with the techniques that have allowed us to understand the genetic history of maize, we are now presented with the challenge of charting its genetic future. The farmer-breeders have been joined by scientists, and acute observation of the phenotypic characteristics of the field crop will be married to the comprehensive charting of the hidden genetic structure of the species.

The story of twentieth century maize improvement has been based on the 'one gene–one function' paradigm, and where molecular engineering of agronomic traits was achieved it was in a step-wise manner where one new allele or one novel transgene at a time were introduced into relevant backgrounds. A schematic representation of all the diverse strands of modern genomics-related research that impact on maize improvement is shown in Figure 6.4.

The paradigm represented in Figure 6.4 is of course, just one way to view the current situation in maize breeding. It focuses mainly on selection of naturally occurring allelic variation and predicts a continuing central role for selection among classically recombined breeding combinations. If one wishes to employ a more futuristic vision where genetic engineering will play an even greater role in the process of molecular breeding, then targeted gene replacement (e.g. Zhu *et al*. 2000) will have to be mastered, since adding randomly integrated transgenes to the existing gene complement allows mainly for dominant effects, while in practice many applications will require the expression of recessive mutations in a null background. For extensive remodelling of the genetic architecture of the maize genome – for instance, the addition of entire heterologous biosynthetic pathways from exotic sources, an approach which

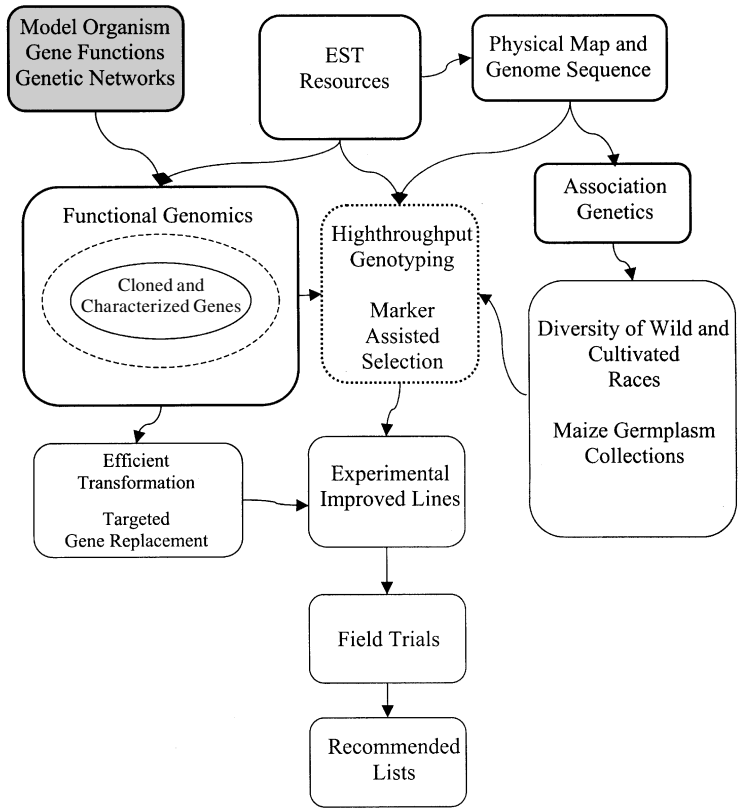


Figure 6.4 A new paradigm for molecular breeding of maize. Four areas of genomics (bold outline) can be undertaken separately but when complete will be integrated into fully cross-referenced information, which can readily be used as a basis for a molecular selection strategy (dashed outline). A comprehensive EST resource would be used both as a template for genome-wide functional analyses as well as providing the identity of a majority of unique maize genes and probes to place each gene on the BAC-based physical map. The ensemble of functional genomics approaches (including transposon mutagenesis and transcriptomic/proteomic/metabolomic data) seeded by a core of known and well-characterized genes and informed by knowledge gained from the rice and *Arabidopsis* models would lead to an exponential increase in the number of molecular targets which could be chosen for selection. Expanded and genotyped ex-situ or in-situ germplasm collections would provide gene diversity maps which could be functionally annotated through genome-wide association studies. Appropriate alleles identified by association genetics would be introgressed into elite and locally adapted varieties assisted by high-throughput, marker-assisted selection followed by phenotypic evaluation of the experimental improved lines thus produced. Transformation provides an alternative route for rapid introduction of novel alleles where appropriate (e.g. non-maize sequences). Improved lines will undergo field trials and the best will (as always) appear on official recommended lists. Background key: the light grey background denotes information obtained through non-maize comparative/model organism approaches; boxes on a clear background denote approaches implemented in maize.

will be required for certain pharmaceutical and nutriceutical applications – maize artificial chromosome technology will need to be developed, based of course on a much better understanding of how native maize chromosomes replicate. These areas which are currently being actively explored, but which are beyond the scope of the present review, allow a tiny glimpse of the next paradigm to follow.

References

Ahn, S. & Tanksley, S.D. (1993) Comparative linkage maps of the rice and maize genomes. *Proceedings of the National Academy of Sciences, USA*, **90**, 7980–7984.

Ajmone-Marsan, P., Castiglioni, P., Fusari, F., Kuiper, M. & Motto, M. (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theoretical and Applied Genetics*, **96**, 219–227.

Anderson, M.D., Prasad, T.K. & Stewart, C.R. (1995) Changes in isozyme profiles of catalase, peroxidase and glutathione reductase during acclimation to chilling in mesocotyls of maize seedlings. *Plant Physiology*, **109**, 1247–1257.

Avramova, Z., Tikhinov, A., SanMiguel, P., *et al*. (1996) Gene identification in a complex chromosomal continuum by local genomic cross-referencing. *Plant Journal*, **10**, 1163–1168.

Batley, J., Mogg, R. & Edwards, K.J. (2002) High-throughput SNuPE assay for genotyping SNPs in the flanking sequence of maize microsatellites. *Molecular Breeding*, **11**, 111–120.

Bellon, M.R. (2001) Participatory research methods for technology evaluation: a manual for scientists working with farmers. Report. Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT).

Bennetzen, J.L. & Freeling, M. (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. *Trends in Genetics*, **9**, 259–261.

Bennetzen, J.L., Chandler, V.L. & Schnable, P.S. (2001) National Science Foundation-sponsored workshop report. Maize genome sequencing project. *Plant Physiology*, **127**, 1572–1578.

Bertoia, L.M., Burak, R. & Torrecillas, M. (2002) Identifying inbred lines capable of improving ear and stover yield and quality of superior silage maize hybrids. *Crop Science*, **42**, 365–372.

Bishop, D.G. (1986) Chilling sensitivity in higher plants: the role of phosphatidylglycerol. *Plant Cell Environment*, **9**, 613–616.

Bressani, R. (1990) Chemistry, technology and nutritive value of maize tortillas. *Food Reviews*, **6**, 225–264.

Brookes, A.J. (1999) The essence of SNPs. *Gene*, **234**, 177–186.

Bureau, T.E. & Wessler, S.R. (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell*, **4**, 1283–1294.

Casa, A.M., Brouwer, C., Nagel, A., *et al*. (2000) Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. *Proceedings of the National Academy of Sciences, USA*, **97**, 10083–10089.

Chory, J., Ecker, J.R., Briggs, S. *et al*. (2000) National Science Foundation-Sponsored Workshop Report: 'The 2010 Project' functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiology*, **123**, 423–426.

Coe, E., Cone, K., McMullen, M., *et al*. (2002) Access to the maize genome: an integrated physical and genetic map. *Plant Physiology*, **128**, 9–12.

Creech, R.G. (1965) Genetic control of carbohydrate synthesis in maize endosperm. *Genetics*, **52**, 1175–1186.

Creencia, R.P. & Bramlage, W.J. (1971) Reversibility of chilling injury to corn seedlings. *Plant Physiology,* **47**, 389–392.

Das, L. & Martienssen, R. (1995) Site-selected transposon mutagenesis at the *Hcf106* locus in maize. *Plant Cell*, **7**, 287–294.

Davis, G.L., McMullen, M.D., Baysdorfer, C., *et al*. (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics*, **152**, 1137–1172.

Desmond, A. & Moore, J. (1991) *Darwin: The Life of a Tormented Evolutionist*. Norton, London.

Devos, K.M., Millan, T. & Gale, M.D. (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. *Theoretical and Applied Genetics*, **85**, 784–792.

Dinges, J.R., Colleoni, C., Myers, A.M. & James, M.G. (2001) Molecular structure of three mutations at the maize *sugary1* locus and their allele-specific phenotypic effects. *Plant Physiology*, **125**, 1406–1418.

Doebley, J. (2001) Meeting report: an evolutionist visits the 43rd Annual Maize Genetics Conference. *Plant Cell*, **13**, 1255–1259.

Doebley, J.F., Durbin, M., Golenberg, E.M., Clegg, M.T. & Ma, D.P. (1990a) Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae). *Evolution*, **44**, 1097–1108.

Doebley, J., Stec, A., Wendel, J. & Edwards, M. (1990b) Genetic and morphological analysis of a maize-teosinte F_2 population: implications for the origin of maize. *Proceedings of the National Academy of Sciences, USA*, **87**, 9888–9892.

Doebley, J., Stec, A. & Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488.

Dorweiler, J., Stec, A., Kermicle, J. & Doebley, J. (1993) *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. *Science*, **262**, 233–235.

Draye, X., Lin, Y.R., Qian, X.Y., *et al*. (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. *Plant Physiology*, **125**, 1325–1341.

Duvick, D.N. (2001) Biotechnology in the 1930s: the development of hybrid maize. *Nature Genetics*, **2**, 71–74.

Duvick, D.N. & Noble, S.W. (1978). Current and future use of cytoplasmic male sterility for hybrid seed production. In: *Maize Breeding and Genetics* (ed. D.B. Walden), pp. 265–277. John Wiley & Sons, New York.

East, E.M. (1936) Heterosis. *Genetics*, **21**, 375–397.

Eathington, S.R., Dudley, J.W. & Rufener, G.K. (1997) Usefulness of marker-QTL associations in early generation selection. *Crop Science*, **37**, 1686–1693.

Edwards, D., Stevenson, D., Forsyth, A., *et al*. (2002) Identification of transposon-tagged maize genes displaying homology to arrayed cDNA clones using *Mutator* insertion display. *Genome Letters*, **1**, 48–55.

Edwards, K.J., Thompson, H., Edwards, D., *et al*. (1992) Construction and characterisation of a yeast artificial chromosome library containing three haploid maize genome equivalents. *Plant Molecular Biology*, **19**, 299–308.

Edwards, K.J., Veuskens, J., Rawles, H., Daly, A. & Bennetzen, J.L. (1996) Characterisation of four dispersed repetitive DNA sequences from *Zea mays* and their use in constructing contiguous DNA sequences using YAC clones. *Genome*, **39**, 811–817.

FAO (1992) *Maize in Human Nutrition*. Food and Agriculture Organisation of the United Nations, Rome.

FAO (2002) FAOSTAT Statistical database of Agricultural data [online]. Food and Agriculture Organisation, Rome. Available from: http://apps.fao.org/page/collections?subset=agriculture.

Feuillet, C. & Keller, B. (1999) High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences, USA*, **96**, 8265–8270.

Fernandes, J., Brendel, V., Gai, X., *et al*. (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiology,* **128**, 896–910.

Frei, O.M. (2000) Changes in yield physiology of corn as a result of breeding in Northern Europe. *Maydica*, **45**, 173–183.

Frame, B.R., Drayton, P., Bagnall, S., *et al*. (1994) Production of fertile transgenic maize plants by silicon carbide whisker-mediated transformation. *Plant Journal*, **6**, 941–948.

Frame, B.R., Zhang, H., Cocciolone, S.M., *et al*. (2000) Production of transgenic maize from bombarded type II callus: effect of gold particle size and callus morphology on transformation efficiency. *In Vitro Cellular and Developmental Biology-Plant*, **36**, 21–29.

Fu, H. & Dooner, H.K. (2000) A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region. *Genome Research*, **10**, 866–873.

Fu, H., Park, W., Yan, X., Zheng, Z., Shen, B. & Dooner, H.K. (2001) The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. *Proceedings of the National Academy of Sciences, USA*, **98**, 8903–8908.

Gai, X., Lal, S., Xing, L., Brendel, V. & Walbot, V. (2000) Gene discovery using the maize genome database ZmDB. *Nucleic Acids Research*, **28**, 94–96.

Gardner, C.O. (1978) Population improvement in maize. In: *Maize Breeding and Genetics* (ed. D.B. Walden), pp. 207–228. John Wiley & Sons, New York.

Gaut, B.S. & Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA*, **94**, 6809–6814.

Goff, S.A., Ricke, D., Lan, T., *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.

Gould, J.H., Devey, M., Hasegawa, O., Ulian, E.C., Peterson, G. & Smith, R.H. (1991) Transformation of *Zea mays* L., using *Agrobacterium tumefaciens* and the shoot apex. *Plant Physiology*, **95**, 426–434.

Hanley, S., Edwards, D., Stevenson, D., *et al*. (2000) Identification of transposon tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays. Plant Journal*, **22**, 557–566.

Hartings, H., Maddaloni, M., Lazzaroni, N., *et al*. (1989) The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO Journal*, **10**, 2795–2801.

Hulbert, S.H., Richter, T.E., Axtell, J.D. & Bennetzen, J.L. (1990) Genetic mapping and characterisation of sorghum and related crops by means of maize DNA probes. *Proceedings of the National Academy of Sciences, USA*, **87**, 4251–4255.

Hull, F.H. (1945) Recurrent selection for specific combining ability in corn. *Journal of the American Society of Agronomy*, **37**, 134–145.

James, C. (2001) Global Review of Commercialized Transgenic Crops: 2001 ISAAA Briefs No. 24. ISAAA. Ithaca, New York.

James, M.G., Robertson, D.S. & Myers, A.M. (1995) Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. *Plant Cell*, **7**, 417–429.

Jones, D.F. (1918) *The effect of inbreeding and cross-breeding upon development.* Connecticut Agricultural Experimental Station Bulletin 207.

Keith, C.S., Buckler, C.S., Hoang, D.O., *et al*. (1993) Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiology*, **101**, 329–332.

Kemper, E.L., Neto, G.G., Papes, F., Moraes, K.C., Leite, A. & Arruda, P. (1999) The role of *opaque2* in the control of lysine-degrading activities in developing maize endosperm. *Plant Cell*, **11**, 1981–1994.

Lin, Y.R., Draye, X., Qian, X., *et al*. (2000) Locus-specific contig assembly in highly-duplicated genomes, using the BAC-RF method. *Nucleic Acids Research*, **28**, E23.

Mangelsdorf, P.C. & Reeves, R.G. (1938) The origin of maize. *Proceedings of the National Academy of Sciences, USA*, **24**, 303–312.

Marton, L.C. (2000) Studies on the inheritance of chilling tolerance in a systematic genetic series of maize (*Zea mays* L.) seedlings in a temperature gradient chamber. *Novenytermeles*, **49**, 475–485.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, J.G., Buckler, E. & Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences, USA*, **99**, 6080–6084.

Meeley, R.B. & Briggs, S.P. (1995) Reverse genetics for maize. *Maize Genetics Newsletter*, **69**, 67–82.

Meghi, M.R., Dudley, J.W., Lambert, R.J. & Sprague, G.F. (1984) Inbreeding depression, inbred and hybrid grain yields, and other traits of maize genotypes representing three eras. *Crop Sciences*, **24**, 545–549.

Mertz, E.T., Bates, L.S. & Nelson, O.E. (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, **145**, 279–280.

Meyerowitz, E.M. (2002) Plants compared to animals: the broadest comparative study of development. *Science*, **295**, 1482–1485.

Mogg, R., Batley, J., Hanley, S., Edwards, D., O'Sullivan, H. & Edwards, K.J. (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theoretical and Applied Genetics*, **105**, 532–543.

Moll, R.H., Longquist, J.H., Fortuna, J.V. & Johnson, E.C. (1965) The relationship of heterosis and genetic divergence in maize. *Genetics*, **52**, 139–144.

Moore, G., Devos, K.M., Wang, Z. & Gale, M.D. (1995) Grasses, line up and form a circle. *Current Biology*, **5**, 737–739.

O'Sullivan, D.M., Ripoll, P.J., Rodgers, M. & Edwards, K.J. (2001) A maize BAC library from the European flint line F2. *Theoretical and Applied Genetics*, **103**, 425–432.

Pejic, I., Ajmone-Marsan, P., Morgante, M., *et al.* (1998) Comparative analysis of genetic similarity among maize lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theoretical and Applied Genetics*, **97**, 1248–1255.

Rabinowicz, P.D., Schutz, K., Dedhia, N., *et al.* (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genetics*, **23**, 305–308.

Rhoades, M.M. (1931) Cytoplasmic inheritance of male sterility in *Zea mays*. *Journal of Genetics*, **27**, 71–93.

Sandstead, R.M., Hites, B.H. & Schroeder, H. (1968) Genetic variations in maize: effects on the properties of the starches. *Cereal Science Today*, **13**, 82–94.

SanMiguel, P., Tikhinov, A., Jin, Y.K., *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.

SanMiguel, P., Gaut, B.S., Tikhinov, A., Nakajima, Y. & Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize. *Nature Genetics*, **20**, 43–45.

Schull, G.H. (1908) The composition of a field of maize. *American Breeder's Association Report*, **4**, 296–301.

Schull, G.H. (1909) A pure-line method of corn breeding. *American Breeder's Association Report*, **5**, 51–59.

Sharopova, N., McMullen, M.D., Schultz, L., *et al.* (2002) Development and mapping of SSR markers for maize. *Plant Molecular Biology*, **48**, 463–481.

Sheridan, W.F. (1982) *Maize for Biological Research*. Plant Molecular Biology Association, Charlottesville, USA.

Shure, M., Wessler, S. & Fedoroff, N. (1983) Molecular identification and isolation of the *Waxy* locus in maize. *Cell*, **35**, 225–233.

Smith, S. (1998) Cultivar identification and varietal protection. In: *DNA Markers* (eds G. Caetano-Anolles & P.M. Gresshoff), pp. 283–299. Wiley-VCH, New York.

Thornsberry, J.M., Goodman, M.M., Doebley, J.F., Kresovich, S., Nielsen, D. & Buckler, E. (2001) *Dwarf8* polymorphisms associated with variations in flowering time. *Nature Genetics*, **28**, 286–289.

Tomkins, J.P., Davis, G., Main, D., *et al*. (2002). Construction and characterization of a deep-coverage Bacterial Artificial Chromosome library for maize. *Crop Science*, **42**, 928–933.

Tsaftaris, A.S. (1996) *Genetics, Biotechnology and Breeding of Maize and Sorghum*. Royal Society of Chemistry, Cambridge, UK.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. & Schmidt, R.J. (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator *Opaque2*. *Proceedings of the National Academy of Sciences, USA*, **94**, 7685–7690.

Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) The limits of selection during maize domestication. *Nature*, **398**, 236–239.

Wessler, S.R., Bureau, T.E. & White, S.E. (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetic Development*, **5**, 814–821.

Yu, J., Songnian, H., Wang, J., *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. spp. *indica*). *Science*, **296**, 79–92.

Yunes, J.A., Cord Neto, G., da Silva, M.J., Leite, A., Ottoboni, L.M. & Arruda, P. (1994) The transcriptional activator *Opaque2* recognizes two different target sequences in the 22-kD-like alpha-prolamin genes. *Plant Cell*, **6**, 237–249.

Zhang, Q., Arbuckle, J. & Wessler, S.R. (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proceedings of the National Academy of Sciences, USA*, **97**, 1160–1165.

Zhu, T., Mettenburg, K., Peterson, D.J., Tagliani, L. & Baszczynski, C.L. (2000) Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. *Nature Biotechnology*, **18**, 555–558.

7 Plant genomics and its impact on wheat breeding

James A. Anderson

7.1 Introduction

For more than a century, wheat breeders have successfully manipulated the wheat genome without knowledge of the chromosomal location or effect for most of the genes important in its production and use. Impressive gains in yield were achieved during the second half of the twentieth century, with annual increases averaging 1.8% on a worldwide basis (Aquino *et al.* 1999). Despite some recent concerns that a yield plateau may be on the horizon, improvements in disease resistance and end-use quality continue (Rajaram 1999). Breeding can be described as the reshuffling of genes and linkage blocks, followed by the selection of desirable progeny. In the case of wheat, and most other plants, this means testing hundreds to thousands of new progeny (genetic combinations) in breeding nurseries each year. This is a process largely of trial and error, in which the chances of producing superior new genetic combinations are improved by extensive knowledge of germplasm attributes and experience as to which parental combinations are likely to yield superior progeny.

Some of the issues faced by wheat breeders include: choice of parents, identification and utilization of beneficial genes and alleles; identifying superior genetic combinations, selection procedures to effectively manipulate quantitatively inherited traits, maintaining or increasing genetic diversity, understanding genotype × environment interaction, and allocation of finite resources for selection and evaluation. The fields of biotechnology (and now more specifically, genomics) have promised improvements aimed at simplifying many of these issues. The purpose of this chapter is to review the current status of wheat genomics and its impact on wheat breeding to date and outline potential future impacts. Special emphasis is placed on the role of markers in wheat breeding.

7.2 Overview of genomics resources in wheat

Common wheat (*Triticum aestivum* L. em thell.) is a disomic hexaploid ($2n = 6x = 42$) containing genomes designated as A, B and D. The diploid progenitors of these genomes have been established as *T. urartu* for the A genome (Dvorak *et al.* 1993), *Aegilops tauschii* for the D genome (McFadden & Sears 1946) and *A. speltoides*, or an extant species, for the B genome (reviewed in Feldman *et al.* 1995). Durum wheat (*T. turgidum* var. *durum*) ($2n = 4x = 28$) contains the A and B genomes.

Numerous other alien species have contributed to wheat improvement (Jiang *et al.* 1994).

7.2.1 Genetic stocks

The polyploid nature of wheat and its tolerance of various forms of aneuploidy have been skilfully exploited for genetic investigations. The mostly widely used genetic stocks are those that are missing either:

- a single chromosome [monosomic (Sears 1954)];
- a pair of chromosomes from one genome, compensated by an extra pair of homoeologous chromosomes [nullisomic-tetrasomic (Sears 1966)];
- an arm [ditelosomics (Sears & Sears 1979)]; or
- sub-arm chromosome pieces (Endo & Gill 1996).

These stocks are illustrated in Figure 7.1. Although chromosomal locations using these stocks are not as precise as genetic linkage map locations, they have the advantage of requiring only intergenomic rather than intragenomic polymorphism. An impressive array of genes has been located to chromosomes in wheat, thanks to the widespread availability of these aneuploid stocks, most of which have been

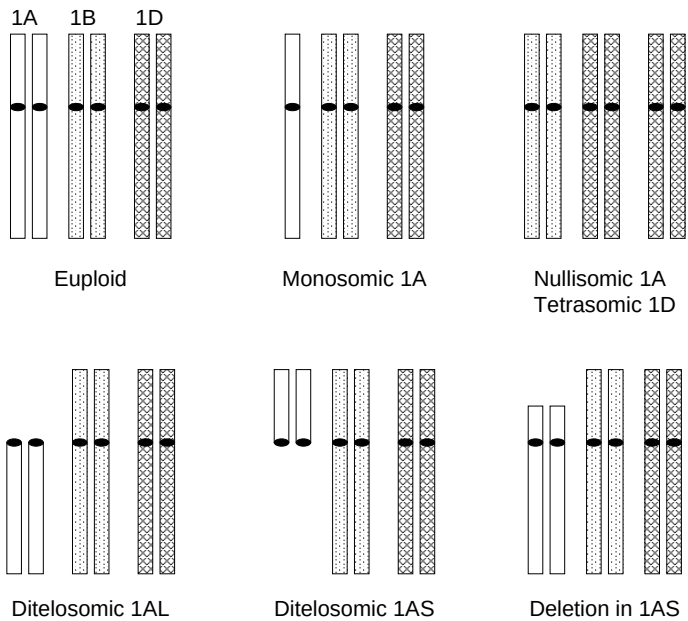


Figure 7.1 Examples of aneuploid stocks in wheat. Common wheat contains seven homeologous groups. One example of each type of aneuploid stock is shown, using homeologous group 1 chromosomes as an example.

developed in the background of 'Chinese Spring'. The *Catalogue of Gene Symbols for Wheat* (McIntosh *et al.* 1998) lists all these locations and is updated annually at http://www.ksu.edu/wgrc/Publications/AWN/AWN.html.

7.2.2 Genetic and physical maps

During the 1990s, molecular mapping in wheat lagged behind many other crops due to its large genome (Arumeganathan & Earle 1991) and low levels of genetic polymorphism (Anderson *et al.* 1993). Wheat has been estimated to have a genome more than six-fold larger than maize, and more than 30-fold larger than rice, due largely to high amounts of repetitive DNA (~80%) (Smith & Flavell 1975). During the past decade, linkage maps containing hundreds of DNA markers each have been produced (Langridge *et al.* 2001). The most widely used and referenced linkage map was developed in a recombinant inbred population from the cross of 'Opata' and a synthetic hexaploid (Nelson *et al.* 1995a, b, c; Van Deynze *et al.* 1995b; Marino *et al.* 1996). The synthetic hexaploid was formed by crossing the durum cultivar 'Altar 84' (AB genomes) with *Aegilops tauschii* (accession 219) (D genome) followed by colchicine doubling (Nelson *et al.* 1995c). The greater genetic diversity between these parents resulted in a higher rate of marker polymorphism compared with potential intraspecies crosses (Anderson *et al.* 1993). The maps were produced largely with restriction fragment length polymorphisms (RFLPs) based on both cDNA and genomic clones, and have been supplemented in some cases with simple sequence repeats (SSRs) (Röder *et al.* 1998a, b; Pestsova *et al.* 2000a) and amplified fragment length polymorphisms (AFLPs) (Lotti *et al.* 2000; Peng *et al.* 2000a; Chalmers *et al.* 2001).

The development of the wheat deletion lines using C-bands as reference markers facilitated the cytologically based mapping of many of these markers. Physical maps have been published for all seven groups (Werner *et al.* 1992; Gill *et al.* 1993, 1996a, b; Kota *et al.* 1993; Hohmann *et al.* 1994; Delaney *et al.* 1995a, b; Mickelsonyoung *et al.* 1995; Weng *et al.* 2000). The correspondence between genetic and physical maps in a region on chromosome 3BS that contains a QTL for *Fusarium* head blight resistance is shown in Figure 7.2.

Large-insert libraries have been developed to facilitate genomics activities. Bacterial artificial chromosome (BAC) libraries have been produced for *T. monococcum* (A genome) (Lijavetzky *et al.* 1999), *A. tauschii* (D genome) (Moullet *et al.* 1999), and libraries are in development for *T. durum*, cultivar 'Langdon' (AB genomes) (J. Dubcovsky, pers. commun.) and *T. aestivum* cultivar 'Chinese Spring' (ABD genomes) (Chalhoub *et al.* 2002). Also, three large-insert libraries of Chinese Spring were prepared in a transformation-competent artificial chromosome (TAC) vector (Liu *et al.* 2000). The status of these and other libraries in development can be found at http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/ITMIbac/ITMIBAC.htm

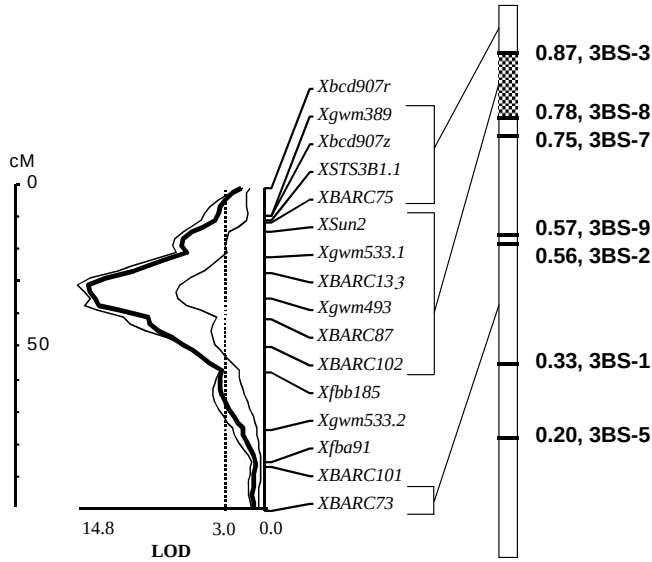


Figure 7.2 Genetic (left) and cytologically based physical (right) maps of chromosome 3BS for *Fusarium* head blight (FHB) resistance in an RIL population from the cross Sumai 3/Stoa. The thick line contour in the genetic map is based on the mean of the two greenhouse FHB evaluation experiments, and the thin line contours are based on individual experiments (Anderson *et al.* 2001). The breakpoints of deletion lines are marked by lines across the physical map. The fraction length (FL) of the retained arm and the name of the deletion lines are listed on the right of the chromosome. The patterned deletion bin is the most likely chromosome fragment that contains *Qfhs.ndsu-3BS*.

7.2.3 Gene clustering and density

Comparison of the location of cDNA and genomic clones on the genetic and cytologically based physical maps of wheat demonstrated a poor correspondence between genetic and physical distance. There is a strong correlation between the presence of genes and rate of recombination (Gill *et al.* 1996a, b; Faris *et al.* 2000; Sandhu *et al.* 2001; Sandhu & Gill 2002). The short arm of group 1 chromosomes is particularly well studied (Gill *et al.* 1996b; Sandhu *et al.* 2001). This arm contains genes coding for grain storage proteins, male sterility and fertility restoration, and resistance to leaf and stem rust, powdery mildew and tan spot. Approximately 75% of the marker loci on chromosome 1BS are concentrated into two gene-rich regions: one representing 6% and 8% of the total length of the chromosome arm and 82% and 17% of the recombination, respectively (Gill *et al.*, 1996b; Sandhu *et al.*, 2001). The remaining 1% of the recombination was contained in the other 86% of the chromosome. The long arm of group 5 chromosomes contains five gene-rich regions (Gill *et al.* 1996a; Faris *et al.* 2000). One region in particular represented 4% of the chromosome arm, yet contained 55% of the markers.

Estimates of base pairs per cM – one estimate of gene density – varies dramatically in wheat, depending upon the chromosomal location. Estimates of the kb/cM ratio of group 1 chromosomes varies from 159 in the terminal 7 Mbp to >5000 in regions near the centromere (Sandhu & Gill 2002). A physical/genetic distance ratio of 20 kb/cM was observed in a BAC clone of *A. tauschii* that contains seed storage protein loci (Spielmeyer *et al.* 2000a). This region is believed to be a recombination hot spot. In a 13.8-kb wheat genomic fragment containing the *Lrk10* gene, maximal gene density was one gene per 4 to 5 kb (Feuillet & Keller 1999). Although these few estimates of gene density in wheat vary depending on the genome region, it is encouraging that the maximum gene density is similar to that observed in *Arabidopsis* (*Arabidopsis* Genome Initiative 2000) and rice (Goff *et al.* 2002; Yu *et al.* 2002).

7.2.4 Synteny and comparative mapping

With the exception of the well-documented 4AL-5AL-7BS translocation (Naranjo *et al.* 1987; Anderson *et al.* 1992; Devos *et al.* 1995; Nelson *et al.* 1995c), an inversion in chromosome 4A (Devos *et al.*, 1995; Nelson *et al.*, 1995c), and a small translocation between 2BS and 6BS (Devos *et al.*, 1993), the three genomes of common wheat are completely syntenous at the resolution of current maps. Moreover, this synteny has been extended to the A and D genome diploid progenitors as well as barley (Kam-Morgan *et al.* 1989; Van Deynze *et al.* 1995b; Dubcovsky *et al.* 1996). Several examples of functional genes being located in syntenous regions in wheat relatives have also emerged, including: aluminium tolerance, *AltBH* on 4DL (Riede & Anderson 1996) with *Alp* on 4HL in barley (Tang *et al.* 2000); seed dormancy on group 4 chromosomes of wheat and barley (Kato *et al.* 2001); seed storage proteins in diploid, tetraploid and hexaploid wheat (Dubcovsky *et al.* 1997); and the vernalization gene *Vrn-A1* in *T. monococcum* and *T. aestivum* (Dubcovsky *et al.*, 1998). The complete synteny with its diploid relatives will undoubtedly have positive implications for genomic investigation of wheat. The main advantage of working at the diploid level is a decrease in the level of duplicated sequence, making navigation and interpretation easier.

The presence of syntenous regions in the genomes of rice, maize, and wheat was demonstrated using RFLP probes and morphological markers (Ahn *et al.* 1993). A common set of anchor probes – RFLPs that hybridize well to all of the major cereal crops – was developed (Van Deynze *et al.* 1998) and has been extensively used as reference markers. At the sub-chromosome arm level, the synteny of wheat has been resolved relative to several other grass species including maize, rice, oat, barley, rye, pearl millet and foxtail millet (Moore *et al.* 1995; Van Deynze *et al.* 1995a; Devos *et al.* 1998, 2000). Investigations of syntenous regions in greater detail has revealed rearrangements and/or regions of apparent non-homoeology (Kojima & Ogihara 1998; Sarma *et al.* 2000).

7.2.5 Sequence-based homology

The preceding results were based on Southern hybridization-based homology. Today, comparative genomics is moving forward based on sequence homology in addition to hybridization-based homology. The completion and availability of genomic sequences of the small genome model species *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative 2000) and rice (*Oryza sativa* L.) (Goff *et al.* 2002; Yu *et al.* 2002) will greatly facilitate these efforts. One current measure of the genomics status of a species is its number of expressed sequence tags (ESTs) deposited in the public databases. At the time of writing, wheat has more ESTs (>40 000) deposited in GenBank than any other plant species. The majority of the wheat ESTs were provided by three groups: participants in the NSF-funded US wheat EST project entitled 'The structure and function of the expressed portion of the wheat genomes' (http://wheat.pw.usda.gov/NSF/); Y. Ogihara and K. Murai at the National Institute of Genetics, Shizuoka, Japan; and Dupont. One goal of the US EST project is to map 10 000 ESTs to a sub-arm chromosome location using the deletion lines. Sequence-based comparisons are facilitated by web-accessible databases and bioinformatic tools. One such example is the *Gramene* database, located at http://www.gramene.org/.

7.2.6 Will rice genomics contribute to wheat improvement?

The high proportion of repetitive DNA in wheat and extensive homology with smaller genome grasses such as rice led to speculation that the future of map-based cloning of wheat genes was via their small genome relatives. Rice genomics is in a much more advanced state compared with other cereals as a result of its status as the model genetic monocot (Sasaki 2001; Yuan *et al.* 2001; Goff *et al.* 2002; Yu *et al.* 2002). The discovery of function for the ~40 000 genes of rice will be accomplished by a variety of methods including homologies with genes of known function in *Arabidopsis* and other species, cloning based on phenotype, and expression analysis. Even with all these options, it will be some time before the identity and function of all agriculturally relevant genes are revealed. At the very least, the synteny with rice can be exploited as a means of providing markers to more fully saturate the homologous region in wheat. The positions of ESTs of five grass species relative to a rice BAC insert are shown in Figure 7.3. Such bioinformatics tools will facilitate comparative mapping and genomics investigations of a large proportion of ESTs identified in the grasses by leveraging the information from all these species.

If a trait or biochemical pathway exists in both rice and wheat, then rice may serve as the source or intermediary to find the homologous gene in wheat. This should be the case for the majority of genes, though it is likely that only a small fraction will be important targets for selection. Sequence-based comparisons of synteny have been published for wheat group 1. Overall good conservation of gene order and high gene density at gene loci in wheat, barley, maize and rice was found (Feuillet & Keller 1999). However, some rearrangements were detected. Analysis of BACs from barley

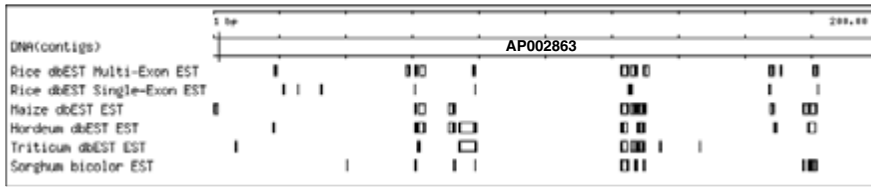


Figure 7.3 A screen shot of Gramene (www.gramene.org/), a database of comparative grass genomics, displaying ESTs of rice, maize, barley, wheat and sorghum that have homology with the rice chromosome 1 BAC clone AP002863.

and rice that each hybridized to a wheat genomic clone showed similar gene content, but some differences in copy number and orientation (Dubcovsky *et al.* 2001).

Although considerable optimism for the exploitation of rice genomics information for wheat improvement is justified, there are several important traits that are unique to wheat, including genes involved in bread-making properties and several important diseases. Rice genomics information may be of limited value in these situations. One example that can illustrate this problem is with disease resistance genes which, in cereal genomes, may have arisen after the divergence of each species (Sasaki 2001). The disease resistance genes that exhibit gene-for-gene interactions, in particular, may be evolving at such a rapid rate that comparative mapping will not be possible. As a case in point, the syntenous region of the barley stem rust resistance gene *Rpg1* was identified in rice, but no candidate genes were identified after sequencing the rice BAC expected to contain the homologue (Han *et al.* 1999). A candidate gene was recently identified from a barley BAC library (Brueggeman *et al.* 2002). In accordance with previous attempts by these authors, no homologue to this candidate gene was found in rice. Another pitfall in the use of rice is that a number of rearrangements exist, even at the sequence level (Feuillet & Keller 1999; Sarma *et al.* 2000). Homology searches also must be carefully examined to take into account the presence of duplicated sequences in each species of interest (Devos & Gale 2000).

7.3 Gene cloning in wheat

An efficient transposon-based tagging system has yet to be developed for wheat. Therefore, cloning activities are concentrated in two areas: cloning based on a gene of similar function; or map-based cloning. Because the function of most agriculturally important genes is not known, map-based cloning is the only recourse. Several recent developments have encouraged map-based cloning efforts in wheat and its direct progenitors. These include:

- the development of BAC libraries;
- rapidly growing EST databases;
- the presence of gene-rich regions; and
- a relatively high gene to kb ratio in the gene-rich regions.

Three examples of cloning agronomically important wheat genes are described below. The reduced height gene, *Rht1*, was cloned based on its homology with a dwarfing gene in *Arabidopsis*; the putative *Cre3* gene was identified as a result of its co-segregation with an NBS-LRR-containing gene family; and *Lrk10* was positionally cloned.

7.3.1 Rht1

One of the most important genes in modern wheat production, the semi-dwarfing gene, *Rht1*, was recently cloned (Peng *et al.* 1999) by taking advantage of its homology to a dwarfing gene (*GAI*) in *Arabidopsis* that had been previously cloned and characterized (Peng *et al.* 1997). The sequence of *GAI* was used to find a homologue in rice by in-silico homology. Primers based on the rice homologue were used to amplify (by PCR) the homologous regions of wheat (*Rht-B1* and *Rht-D1*) and maize (*d8*). All the grass homologues mapped to syntenous regions as indicated by the locations of other cross-hybridizing genomic clones. The reduced height gene *Dwf2* in barley was previously shown to be homologous with *Rht-B1*/*Rht-D1* in wheat (Ivandic *et al.* 1999).

7.3.2 Cre3

The *Cre3* gene that confers a high level of resistance to the cereal cyst nematode (CCN) (*Heterodera avenae*) in wheat was introgressed from *A. tauschii* (Eastwood *et al.* 1991, 1994). A gene family predicted to encode a nucleotide binding site and leucine-rich repeat (NBS-LRR) protein was found to co-segregate with the *Cre3* locus (Lagudah *et al.* 1997). This gene was physically mapped to the distal 0.06 fragment of the long arm of wheat chromosome 2D. The segregation of other members of this family corresponds to the location of other CCN resistance genes that are in non-syntenic locations (Lagudah *et al.* 1997, 1998).

*7.3.3 Leaf rust (*Lr10*)*

The leaf rust resistance gene *Lr10* is derived from CIMMYT germplasm and is a component of multigene resistance of many North American wheat cultivars (Pretorius & Roelfs 1996). A putative resistance gene (*Lrk10*) at the same genetic locus as *Lr10* has been cloned (Feuillet *et al.* 1997; Stein *et al.* 2000) and encodes a receptor-like protein kinase.

7.4 Transgenics

Fertile transgenic plants of common wheat were first reported in 1993 (Weeks *et al.* 1993), and reports of success in other laboratories followed (Becker *et al.* 1994; Nehra *et al.* 1994) such that durum also has been transformed (Bommineni *et al.* 1997; He *et al.* 1999). Despite one report of successful transformation of wheat with *Agrobacterium tumefaciens* (Cheng *et al.*, 1997), particle bombardment of immature embryos remains the most widely used method of transgene delivery. Transformation of wheat is very difficult, owing to poor regeneration of plants from immature embryos of most genotypes. Even the most regenerable genotypes have resulted in transformation efficiencies (fertile transgenics per bombarded embryo) of 0.1 to 2% (Lorz *et al.* 1998; G. Muehlbauer, pers. commun.). This is far below the efficiencies achieved in *Arabidopsis*, which is easily transformed by *A. tumefaciens.* This low efficiency is a bottleneck today for applied purposes, but also will delay efforts to verify the identity of cloned genes via complementation analysis.

No transgenic wheat cultivars are currently in production. However, many are in various stages of field-testing. Target traits include herbicide resistance, pathogen and insect resistance, and grain end-use quality modification. The *CP4* gene that gives resistance to the herbicide glyphosate has been successfully inserted into common wheat (Zhou *et al.* 1995). The genes being inserted for disease resistance include chitinases, glucanases, thaumatin-like proteins that have been cloned from other plants, as well as genes from the pathogen itself. Initial reports of the functional properties of the transgenic wheats have been largely encouraging. Resistance to powdery mildew (Bliffeld *et al.* 1999; Oldach *et al.* 2001); stinking smut (Clausen *et al.* 2000); delay in onset of symptoms of *Fusarium* head blight (FHB) (Chen *et al.* 1999); increased resistance to wheat streak mosaic virus (WSMV) (Sivamani *et al.* 2000) and an absence of barley streak mosaic virus (BSMV) symptoms (Zhang *et al.* 2001) have been reported. Resistance to some insects has also been achieved using transgenic approaches; these include resistance to Angoumois grain moth (*Sitotroga cerealella*) (Altpeter *et al.* 1999) and to a grain aphid, *Sitobion avenae* (Stoger *et al.*, 1999).

Because many of the genes being inserted for pathogen and insect resistance are general defence response genes that have been cloned from other species, the resulting transgenic plants should be tested against as broad a spectrum of pests as possible. Although the transgenic event may not give resistance to the intended target, it may provide defence against other pests. From the examples cited above, partial resistance or reduction of disease symptoms is more common than complete resistance or immunity. This will make these transgenics less appealing for breeders, though these genes – when combined with native host plant resistance genes – may help combat several of our more difficult pests.

Modification of grain end-use quality has been accomplished via transgenic approaches. Both starch properties and the high-molecular weight (HMW) glutenin genes have been targeted to improve or alter the functional properties of flour. The progress and potential of these approaches were recently reviewed (Vasil & Ander-

son 1997; Blechl *et al.* 1998; Shewry *et al.* 2001). Results have included unintended silencing of HMW glutenin subunits or altered expression (Demeke *et al.* 1999; Alvarez *et al.* 2000); abnormal mixing properties (Popineau *et al.* 2001); increased dough strength (Rooke *et al.* 1999; Alvarez *et al.* 2001); and improved baking performance with no obvious deleterious characteristics (Vasil *et al.* 2001).

Transgenic maize and soybean cultivars have been in widespread production for several years, and it is expected that wheat cultivars with unique traits that can only be obtained via transgenic approaches will also be accepted. Additional improvements in the production of transgenic lines using adapted germplasm and more exact developmental and tissue-specific transgene expression will be essential to make this an attractive means of wheat improvement. Given the recent progress in gene identification and cloning in wheat and other plant species, it does not appear as though availability of target genes for introduction or manipulation will be a significant bottleneck.

7.5 Applications/examples of DNA marker technology in wheat breeding

Genomics research has resulted in the identification of more than 150 marker: gene/QTL linkages in wheat (reviewed by Langridge & Chalmers 1998; Gupta *et al.* 1999; Langridge *et al.* 2001). The map locations of these genes are useful for the understanding of genome organization and gene function, and to serve as the starting points for map-based cloning efforts. Additional information for the breeder from these studies includes the number of genes and magnitude of their effects influencing quantitatively inherited traits, as well as usefulness of the marker as a selection tool. Markers also are more generally used as indicators of genetic diversity, without being associated with particular genes. The next two sections will summarize some of the major findings and progress in using markers in wheat to assess genetic diversity and marker-assisted selection.

7.5.1 Assessing genetic diversity

Common wheat is classified according to its growth habit (spring versus winter), grain colour (red versus white), grain hardness, and grain quality attributes. This has provided a form of identity-preserved marketing in many countries where more than one class is grown. Within classes, certain end-use quality characteristics are expected. Most breeding programmes work with only one or two of these classes and confine crossing within a single market class to retain the suite of genes and linkage blocks that are responsible for local adaptation and conformation to the particular market class. Given the high level of intra-class crossing that takes place, there may be a risk of limiting genetic diversity, thus making the crop more vulnerable to changes in disease pressure or races, or limiting long-term improvement. Recent molecular and coefficient of parentage studies have shown similar levels of diversity in modern-day cultivars versus those grown a half-century ago in the UK

(Donini *et al.* 2000), North America (Mercado *et al.* 1996; van Beuningen & Busch 1997), France (Metakovsky & Branlard 1998) and Spain (Metakovsky *et al.* 2000). Diversity that may have been lost was likely represented by genes or linkage blocks that did not provide enhancements in productivity or disease resistance, and this has been compensated by more favourable genes or linkage blocks. The introgression of disease resistance genes from alien species is one way in which both productivity and diversity can be increased simultaneously.

Several studies have attempted to correlate the diversity of parents based on DNA markers or coefficient of parentage with performance of the resulting progeny (Burkhamer *et al.* 1998; Fabrizius *et al.* 1998; Perenzin *et al.* 1998; Bohn *et al.* 1999; Liu *et al.* 1999). The expectation was that greater diversity of parents would result in more genetic variation in progeny and produce a greater proportion of transgressive segregates. Unfortunately, progeny variance was poorly correlated with parental diversity, though in most cases, parents could be logically grouped based on known characteristics. AFLP markers also have been used to effectively assess genetic diversity among wheat cultivars and cluster cultivars based on market class (Barrett & Kidwell 1998; Barrett *et al.* 1998; Bohn *et al.* 1999).

Markers may be valuable for assessing diversity within germplasm collections, prioritizing accessions for evaluation and crossing, and answering questions about the levels of genetic diversity of modern wheats compared to its progenitor species. The analysis of *A. tauschii*, the D genome progenitor of common wheat, will serve as an example for these points. *Aegilops tauschii* has proved to be a highly polymorphic (Pestsova *et al.* 2000b), large reservoir of genes not found in common wheat. Important genes obtained from *A. tauschii* include those providing resistance to Hessian fly (Hatchett & Gill 1981; Cox & Hatchett 1994), cereal cyst nematode (Eastwood *et al.* 1991; Delibes *et al.* 1993), *Septoria* (McKendry & Henke 1994; Loughman *et al.* 2001) and leaf rust (Kerber 1987; Raupp *et al.* 2001). Compared with the A and B genomes, the D genome has shown lower levels of polymorphism in common wheat (Nelson *et al.* 1995b; Röder *et al.* 1998b). One possible reason for this may be that common wheat originated from only one or few original polyploidization events followed by self-pollination. From comparisons of the level of variability in the D genome of *A. tauschii* with that in several common wheats, it appears as though common wheat is the product of multiple polyploidization events (Talbert *et al.* 1998; Lelley *et al.* 2000).

7.5.2 Marker-assisted selection (MAS)

It is clear that of the hundreds of marker:gene/QTL linkages reported in the literature, only a handful have been utilized in MAS programmes. Markers used in Canadian programmes include those for the bunt resistance gene *Bt10* (Demeke *et al.* 1996), high protein gene from *T. dicoccoides* (Mesfin *et al.* 1999; Khan *et al.* 2000), glutenin and gliadin subunits, red/white kernel colour, and the 1BL/1RS translocation (Howes *et al.* 1998). A recent survey of Australian wheat-breeding programmes

indicated that DNA markers are being used to monitor at least 19 genes, or chromosome segments (Eagles *et al.* 2001); however, for only two of the traits – cereal cyst nematode (*Cre1*, *Cre3*) and endosperm waxiness (*Wx-B1*) – more than 600 assays per year were performed on a nationwide basis. A $3.25 million effort to produce new wheat germplasm and cultivars in the US through MAS was recently funded by the USDA's Initiative for Future Agriculture and Food Systems (IFAFS) (http://maswheat.ucdavis.edu/index.htm). The overall goal of this project is to transfer new developments in wheat genomics and biotechnology to wheat production through the organization of a national wheat consortium including 12 wheat breeding and research programmes across the US. Available molecular markers will be used to transfer 23 genes providing resistance to fungi, viruses and insects, and 17 gene variants related to bread, pasta and noodle quality into more than 300 new, adapted cultivars or breeding lines belonging to all major market classes of US wheat. In many cases, these genes will be pyramided with the help of molecular markers.

7.5.2.1 DNA markers for disease resistance genes

The largest class of DNA marker/gene linkages in wheat is for disease resistance genes, and their use in MAS is one example of the application of genomics tools in wheat breeding. Genetic markers for disease resistance genes can increase efficiency of selection and will most likely be useful in cases of:

- low-heritability diseases (e.g. *Fusarium* head blight, cereal cyst nematode);
- pyramiding genes (e.g. rusts, powdery mildew);
- lack of affordable or effective screening methods;
- absence of pathogen (e.g. karnal bunt); and
- accelerated backcrossing to recover the recurrent parent, especially in cases where non-adapted germplasm is used as the gene donor.

The identification of markers for qualitatively inherited genes is straightforward, typically based on F_2, F_3, doubled haploid or recombinant inbred progeny from single crosses. Quantitative traits, on the other hand, require inbred lines or doubled haploids that can be replicated to allow for phenotypic evaluation in multiple environments. QTL have been identified for the wheat diseases of leaf rust (Nelson *et al.* 1997; William *et al.* 1997), karnal bunt (Nelson *et al.* 1998), tan spot (Faris *et al.* 1997; Effertz *et al.* 2001), *Fusarium* head blight (Bai *et al.* 1999; Waldron *et al.* 1999; Anderson *et al.* 2001; Buerstmayr *et al.* 2002), stripe rust (Borner *et al.* 2000; Peng *et al.* 2000b, c; Singh *et al.* 2000; Spielmeyer *et al.* 2000b) and powdery mildew (Chantret *et al.* 2000, 2001).

Although the number of DNA markers for disease resistance genes in wheat increases on a monthly basis, few examples of their use in breeding have emerged. Examples include cereal cyst nematode (*Cre1*, *Cre2*, *Cre3*), eyespot (*Pch1*), *Fusarium* head blight (*Qfhs.nsdu-3BS*), leaf rust (*Lr37*), powdery mildew (*Pm1*), stem rust (*Sr6*, *Sr22* and *Sr38*), stripe rust (*Yr17*) and wheat streak mosaic virus (*Wsm1*). In the

following sections, three examples of marker use in breeding for disease resistance are highlighted: the case of eyespot (*Pch1*), cereal cyst nematode (*Cre1*, *Cre3*), and *Fusarium* head blight (*Qfhs.ndsu-3BS*).

Eyespot (strawbreaker foot rot)

Eyespot (syn. strawbreaker foot rot) is caused by *Tapesia yallundae* Wallwork & Spooner, and results in the weakening of stem bases and lodging. Disease severity of lines is assessed in a quantitative manner following inoculations of wild-type strains in the field or growth chamber evaluation after inoculation with a β-glucuronidase (GUS)-transformed strain (de la Peña & Murray 1994). One potent eyespot resistance gene, *Pch1*, is derived from *Aegilops ventricosa* and was introgressed to a segment on chromosome 7D and is associated with an endopeptidase isoenzyme marker (*Ep-D1b*) (McMillin *et al.* 1986; Worland *et al.* 1988). This marker has been widely used by breeding programmes in the Pacific Northwest of the US and in Europe. A second resistance gene, *Pch2*, from the French cultivar 'Capelle Desprez' also has been mapped (de la Peña *et al.* 1996). This gene does not confer as much resistance as *Pch1* and, therefore, has not been used to date in MAS.

In the USDA-ARS wheat breeding programme at Pullman, WA, lines are screened for presence of the diagnostic *A. ventricosa*-derived isoenzyme after preliminary yield trials (K. Campbell, pers. commun.). Isoenzyme screening is delayed to this stage because large numbers of materials (headrow stage) are more efficiently screened in inoculated field nurseries, and this gene is present in a large proportion of parental materials. One person can process about 60 samples per day, and a total of about 150 lines is tested for the isoenzyme marker on a yearly basis. Due to the presence of a third isoenzyme phenotype that is not associated with resistance or susceptibility, all lines believed to possess *Pch1* are tested under field conditions to confirm their eyespot resistance for at least 2 years prior to cultivar release. The field screening also facilitates identification of germplasm containing genes that provide greater levels of resistance than *Pch1* alone. Certainly, part of the appeal of a marker for this gene is that there is absolute linkage with *Pch1*, owing to its presence on an alien segment that does not recombine with wheat.

Cereal cyst nematode

Damage by the cereal cyst nematode causes 8% annual yield losses in south-eastern Australia (Eastwood *et al.* 1991). The traditional means of screening for resistance to this pathogen is time-consuming, inconsistent and relatively expensive. In fact, the difficulty of the phenotypic screen was the primary motivation for the adoption of markers by the breeders. Molecular markers for three resistance genes, *Cre1* (Williams *et al.*, 1994, 1996), *Cre3* (Eastwood *et al.* 1994; Lagudah *et al.* 1997) and *Cre6* (Ogbonnaya *et al.* 2001a) have been reported. Markers useful in breeding for all three of these genes have been developed and validated (Ogbonnaya *et al.* 2001b). The Victorian Institute for Dryland Agriculture's wheat breeding programme is running about 3000 assays annually to detect the presence of *Cre1* (using an RFLP marker) and *Cre3* (using an allele-specific PCR marker) in their populations (Eagles

et al. 2001). Within Australia, only selection for the waxy gene, *Wx-B1*, is more widely practised using markers than the cereal cyst nematode genes.

Fusarium head blight (FHB)

Fusarium head blight resistance is quantitatively inherited and numerous chromosomal regions have been reported to influence resistance (reviewed by Bai & Shaner 1994). Breeding for resistance to this disease using conventional methods is difficult due to inconsistencies caused by dependence on high moisture and conducive temperatures at flowering time. The Chinese wheat cultivar 'Sumai 3' and its derivatives have been successfully used as FHB resistance sources worldwide. A major QTL, *Qfhs.ndsu-3BS*, derived from Sumai 3 has been identified and verified in several research groups (Bai *et al.* 1999; Waldron *et al.* 1999; Anderson *et al.* 2001; Buerstmayr *et al.* 2002) using molecular marker analysis. Depending on the population, the QTL explains 17 to 60% of the phenotypic variation in FHB resistance. We have used the *Qfhs.ndsu-3BS* QTL in both MAS and for diagnostic purposes in parental selection. In both 2000 and 2001, parents of families suspected to be segregating for this QTL based on pedigree were screened for their DNA marker allele type at three or four SSR loci near the QTL. In 2001, this represented 135 parents of 55 families. As illustrated in Figure 7.4, there is a high level of polymorphism information content (PIC) inherent in the markers in this region. PIC values were very high despite very close relationships of several of the parents. In fact, this set of parents includes 23 full sib lines. The Sumai 3 allele types appear to be relatively unique to this genotype and its relatives; therefore, these markers are predictive of the presence of this QTL in our populations.

We have extended this work to also examine the *Qfhs.ndsu-3BS* haplotypes of other resistance sources (Liu & Anderson 2003). Some 54 FHB-resistant lines from

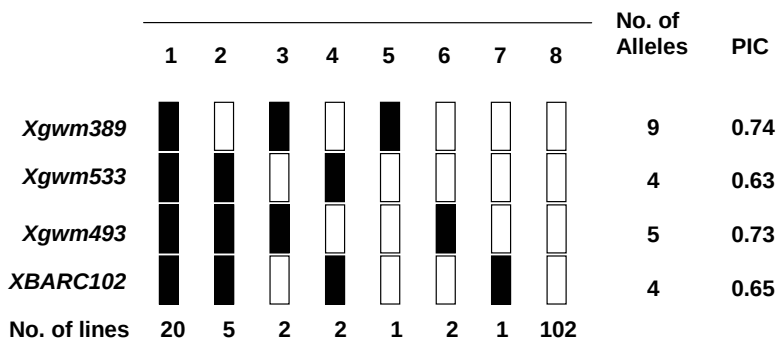


Figure 7.4 Haplotype figures for 135 wheat lines using SSR markers near the *Qfhs.ndsu-3BS* QTL. Darkened boxes indicate the Sumai 3 (resistant) allele at this locus and open boxes indicate a non-Sumai 3 allele. PIC values were calculated using all SSR alleles, instead of characterizing them as Sumai 3 and non-Sumai 3 type, as they are shown in this figure. The wheat lines are parents of 55 populations of F_4 lines that contained at least one parent believed to contain this QTL. From pedigree and FHB reaction, 29 of the 135 lines are believed to contain this QTL.

throughout the world and 20 susceptible lines of historical prominence in North America Spring Wheat improvement were genotyped with five SSR markers. The SSR markers *Xgwm389*, *Xgwm533*, *XBARC133*, *Xgwm493* and *XBARC102* had PIC values 0.84, 0.74, 0.79, 0.71 and 0.70, and detected 14, 12, 8, 9 and 12 alleles, respectively. As found with the breeding lines, the Sumai 3 haplotype is rare. The resistant lines that are distinct from Sumai 3 at this QTL are good candidates for further genetic study and for use as parents. Analysis of other resistance sources in this manner will allow us to prioritize our research and focus on those that do not contain this QTL.

The effectiveness of MAS for this QTL is being assessed using near-isogenic lines (NILs) identified in our regular breeding materials. Using the process shown in Figure 7.5, several dozen NIL pairs have been developed at the *Qfhs.ndsu-3BS* locus by selecting for F_3-derived F_4 lines that are heterozygous for this QTL region. Single, heterozygous plants are identified and self-pollinated. The two homozygous classes are identified from the progeny and seed increased for phenotypic evaluation. If desired, a greater level of inbreeding could be obtained by repeated selection for

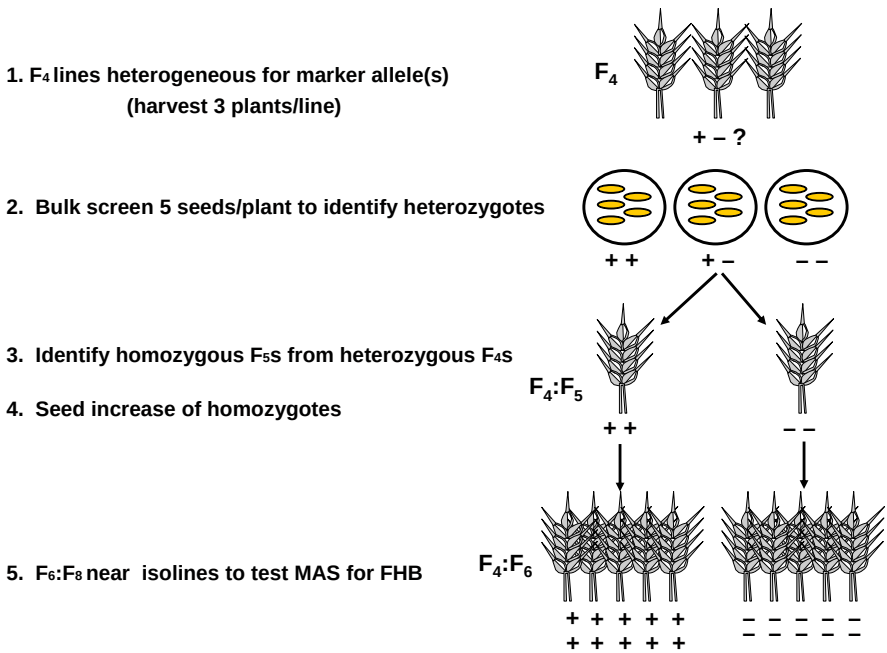


Figure 7.5 Development of QTL-near isogenic lines (QTL-NILs) for testing the effect of QTLs in multiple genetic backgrounds. In step 1, heterogeneous F_4 lines are identified after PCR amplification of marker alleles from a bulk harvest of $F_{3:4}$ plants. When applied to multiple families segregating for the presence of a QTL, the resulting QTL-NILs are used to predict the effectiveness of QTL selection across germplasm backgrounds and in the presence of other QTLs affecting the trait.

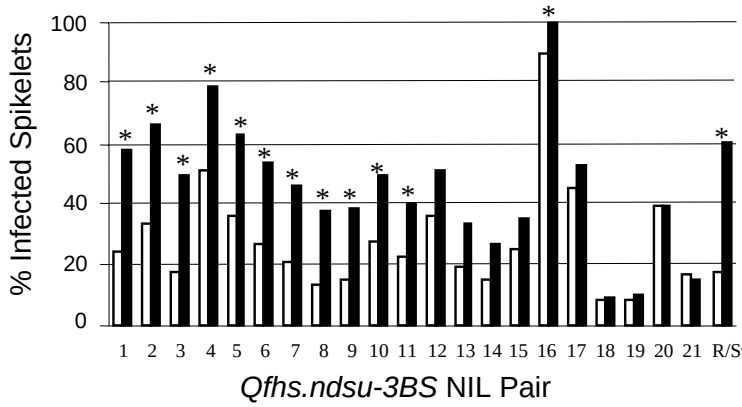


Figure 7.6 Results from greenhouse point-inoculation FHB screening of QTL-NIL pairs. Fifteen of the 21 pairs are from unique single-cross combinations, with one parent having Sumai 3-derived resistance. Open bars indicate lines with Sumai 3-type SSR marker alleles; black bars indicate sib lines without Sumai 3 alleles. [a]Mean of resistant (R) parents with Sumai 3 alleles and susceptible (S) parents. *Significant at $P < 0.05$.

the heterozygote, followed by self-pollination. The FHB reaction of 21 NIL pairs, representing 15 cross-combinations (multiple NIL pairs for some families), is shown in Figure 7.6. There were significantly fewer infected spikelets in the line containing the QTL in 12 out of the 21 pairs. For the remaining pairs, the difference in FHB reaction of the two NILs was very small, and in most cases probably reflects a high level of resistance in the genetic background, presumably due to the presence of other, unselected QTLs. These preliminary data suggest that this QTL significantly increases FHB resistance, regardless of genetic background, and MAS should be continued.

7.5.2.2 Limitations to marker-assisted selection in wheat, and possible solutions

Why have relatively few examples of MAS in wheat breeding emerged? In many breeding programmes, a great deal of research is devoted to optimizing selection strategies for key traits. Therefore, DNA marker technology is often competing with a well-developed conventional screening system. It may also be helpful to point out that the majority of selection – at least in the generations before yield testing – is done in a matter of seconds, by walking through a nursery of thousands of headrow selections or visually examining seed samples. Several criteria must be satisfied for MAS to be considered: efficiency/gain compared with conventional selection; usefulness of markers in identifying the gene or QTL with respect to the elite germplasm; and cost, throughput and expertise required. Each of these criteria, together with an example or approach as to how this may be overcome in wheat, will be outlined in the following sections.

Efficiency/gain compared with conventional selection

The proportion of the variance explained by markers is an important consideration when dealing with quantitative traits. This is not an issue with qualitative traits expressed as distinct phenotypes (e.g. resistant versus susceptible), except to the extent that the genotype does not equate to phenotype (marker) due to recombination. In theory, MAS is most advantageous when heritability is low and the proportion of variance explained by markers is high. However, these situations are not common. A preponderance of QTL studies reveal moderate effects (R^2 of 10–20%) at one or a few loci, and several other loci explaining less than 10% of the variance. Population sizes used to establish QTL linkages (100–200) are not large enough to provide precise estimates of QTL effects (Beavis 1998). Instead, marker effects in such small populations are usually overestimated.

Just as important is how well the alleles tagged by the markers perform when present in other genetic backgrounds. Parents for developing mapping populations are chosen, in part, due to their disparate values for the trait or traits to be mapped. If the QTL is subject to epistatic effects, its effect could be quite different when present in other (adapted) genetic backgrounds. Two common approaches to this issue are:

- assessing the QTL effect in additional mapping populations or validation populations; or
- introducing the QTL into various breeding populations and comparing phenotypic values of progeny selected using markers versus a random sample or conventionally selected progeny.

Unfortunately, both of these approaches are very time-consuming. We have chosen to investigate QTL effects over multiple genetic backgrounds through the development of QTL-NILs, as illustrated in Figure 7.5. Developing NILs over a representative cross-section of the breeding germplasm in this manner fulfils two important functions by assessing:

- the informativeness of markers *per se* for the QTL; and
- the effect of the QTL in different genetic backgrounds, thus taking into account the presence of other, unselected genes.

The resulting genetic material is directly useable in the breeding programme (i.e. the result is not another mapping population that at best may produce new parental material for the breeding programme).

Are the markers useful in your populations?

The marker systems of the late 1980s and early 1990s were plagued by low polymorphism and detection of relatively few alleles (Anderson *et al.* 1993). Microsatellite markers based on SSRs have been the marker of choice in the past 2 to 3 years, and have largely replaced RFLPs, RAPDs and some sequence-tagged site (STS) markers that were originally used to tag genes. Advantages of SSRs include:

Figure 7.7 Thirteen alleles detected with marker *Xgwm389* after amplification of wheat DNA from 74 genotypes. The first and the last lanes are the molecular size ladders.

- no radioactivity;
- small amount of template DNA required;
- amenable to high-throughput procedures;
- PCR primers are easily obtained; and
- high polymorphism information content (PIC) values.

SSR markers have revealed relatively high levels of polymorphism (Plaschke *et al.* 1995). PIC values of 247 SSR loci based on eight varieties ranged from 0.22 to 0.88, with 174 (70%) scoring 0.50 and higher (Harker *et al.* 2001). This research also resulted in the identification of 30 SSR which include loci on all but two chromosome with high PIC values that can be used as a reference set. Such a set of markers could be used as an initial screening step to recover the recurrent parent genome in a backcrossing programme. The high number of alleles that can be detected using SSRs in wheat is shown in Figure 7.7.

The marker information of neighbouring segments can be combined to more clearly discriminate among alleles or haplotypes (see Figure 7.4). The haplotype concept is also used with single nucleotide polymorphisms (SNPs) because it is more informative than analysis based on individual SNPs, and has more power in analysing association with phenotypes (Rafalski 2002). There is some risk of introgressing too large a segment if the markers used are relatively far apart. Ideally, one would want to use a single marker with absolute linkage to the gene (e.g. the gene itself) or flanking markers less than 5 cM apart, or as close as possible, for MAS.

Cost, throughput and expertise required

Wheat breeding programmes operate on small margins compared with other crops. This is especially the case in the US, where the majority of seed planted each year is farmer-saved. As a result, the majority of the wheat breeding effort in the US has

remained in the public sector. Marker screening costs remain relatively high; for example, our costs on a per genotype basis for SSR marker screening are currently about $0.90 per datapoint. The three largest components of this cost are labour for DNA extraction ($0.21), *Taq* polymerase ($0.18), and labour for PCR and gel electrophoresis ($0.12). All of these costs can be substantially reduced by adoption of more high-throughput methods and economies of scale. Marker platforms based on SNPs can eliminate the need for gel electrophoresis (Gupta *et al.* 2001), which would reduce one major source of labour. Cost of reagents, including *Taq* polymerase, can be reduced by multiplexing and reduction of reaction volume through the use of robots or liquid handling systems. Unfortunately, the funding outlay required for this type of equipment is beyond the means of most individual breeding programmes.

Improvements in high-throughput genotyping are being driven by human genomics research to a large extent. SNP markers are well established in human genome activities and are also being adopted by plant researchers (Gupta *et al.* 2001; Martins-Lopes *et al.* 2001). A number of platforms exist for genotyping via SSRs (Nagaoka & Ogihara 1997) and SNPs (Gupta *et al.* 2001).

The switch from RFLP- to PCR-based marker systems has allowed breeding laboratories to become more self-sufficient in mapping and genotyping with markers. This may be a short-lived situation as the high-throughput methods of the future may require higher levels of user expertise and their cost may be too high for many laboratories. One concept for dealing with this situation is the establishment of genotyping centres. These already exist for the Australian wheat breeding programmes (http://www.scu.edu.au/research/cpcg). A similar system was proposed for the US (Van Sanford *et al.* 2001), and the first USDA-ARS funded laboratory is now operational in Manhattan, KS. Analogous to the USDA-ARS regional wheat quality laboratories, the regional genotyping centres will provide support to small grain breeding programmes in the US through use of automated DNA extraction and high-throughput marker screening procedures. They will provide a bioinformatics interface between molecular genetic data and public and private breeding programmes. The regional nature of the genotyping centres should help breeders and mappers focus on valuable traits for their particular production area, market class and production system, to facilitate coordination during peak genotyping periods, and to provide a critical mass of user groups (i.e. breeding programmes) in each of the regions. Concentrating the genotyping activities in regional laboratories will ensure that the expertise and equipment will keep pace with improvements in technology. This allows breeders to focus their attention on choosing populations and markers, and utilization of the resulting information.

7.5.2.3 Beyond MAS: direct allele selection

For most agronomically important genes in wheat, allelic variation has not been well characterized. This is crucial information for the breeder who otherwise has to distinguish alleles based on phenotype. If two alleles produce similar phenotypes such that their effects cannot be statistically discerned, the breeding progress will be

slowed. The identification of a marker for a particular gene is based on its polymorphism and phenotypic difference between parents of a segregating population. One of the criteria for using markers for QTL selection is that the markers explain a sufficient quantity of the phenotypic variation to warrant their use. Because most QTL are not precisely mapped, it is difficult to distinguish a reduced phenotypic effect from recombination between the markers and QTL. This situation can be improved by using the ultimate marker, the gene itself.

Although several years of intensive research are required to positionally clone QTLs, its feasibility has been recently demonstrated. A QTL influencing fruit weight in tomato was obtained by map-based cloning in 2000 (Frary *et al.* 2000). Other QTLs cloned via map-based approaches include the *ED1* flowering time (El-Assal *et al.* 2001) and *FRI* vernalization-sensitive flowering time (Johanson *et al.* 2000) QTL in *Arabidopsis*; and two photoperiod sensitivity genes from rice, *Hd6* (Takahashi *et al.* 2001) and *Hd1* (Yano *et al.* 2000). With the genomics tools that are being developed, cloning of QTL will be easier. In a best-case scenario, markers that are highly specific to the beneficial allele can be developed and utilized in breeding.

The concept of direct allele selection (Sorrells & Wilson 1997) was described as a means of identifying alternative alleles within the sexually compatible gene pool. Alternative alleles are discovered based on sequence differences of a known gene. The development of near-isogenic stocks may be necessary to characterize the phenotypic effects and breeding value of different genes and alleles. The development of near-isogenic lines of 'Thatcher' wheat containing different leaf rust resistance genes has been invaluable in defining the effects of specific genes and in determining the virulence genes contained in different rust races (Kolmer & Dyck 1994). The HMW-glutenin alleles on group 1 chromosomes (*Glu-A1*, *Glu-B1* and *Glu-D1*) and the locus largely responsible for grain hardness, *HA*, are examples of allelic series that are characterized as to their phenotypic effects and the alleles can be readily distinguished based on electrophoretic mobility of DNA fragments or protein.

The HMW-glutenin alleles are known to account for a large portion of the variability in end-use quality among wheat cultivars. These alleles are visualized as different size of protein bands on gels and have been correlated with end-use quality (Payne *et al.* 1987; Lookhart *et al.* 1993; Wang *et al.* 1993; Nakamura 2000). This is valuable information for the breeder to select among the various subunits of the A, B and D genomes to obtain the desired grain quality. In the hard winter wheat growing region of the US, these loci were largely fixed for the desirable alleles for bread-making quality (Lookhart *et al.* 1993), the predominant use of wheat grown in this region. However, this information is still valuable when choosing parents for crossing to be sure that at least one parent contains the beneficial allele. In segregating materials, the HMW-glutenin alleles can be used as a first screen prior to more expensive and time/grain-consuming tests of rheological properties. Knowledge of the specific allele types in cultivated wheat has provided an impetus to discover novel alleles in the relatives of wheat. Glutenin diversity has been examined in the einkorn wheats (Rodriguezquijano *et al.* 1997; Lee *et al.* 1999), *T. dicoccoides* (Pagnotta *et al.* 1995) and

T. turgidum dicoccum (Pfluger *et al.* 2001). As these novel alleles are introgressed into cultivated wheats, their effects on end-use quality can be assessed.

Allelic variation associated with the hardness locus, *Ha*, has been recently characterized (Giroux *et al.* 2000; Martin *et al.* 2001; Morris *et al.* 2001). This gene is located on chromosome 5DS (Symes 1965), and explained more than half the variation in grain hardness when segregating in a cross between hard and soft-textured endosperm wheat (Campbell *et al.* 2001). The putative *Ha* gene has been cloned (Giroux & Morris 1997), although evidence exists that other genes may be responsible for the trait (Osborne *et al.* 2001). Nevertheless, the *PinA* and *PinB* loci are closely associated with the *HA* locus. The sequence analysis of these alleles and subsequent PCR amplification of homologues in other wheat genotypes has revealed additional alleles (Morris *et al.* 2001). Knowledge of the specific allele causing the grain hardness in different germplasm can lead to more efficient manipulation of this trait. In a mapping population of 139 spring wheat recombinant inbred lines (RILs) segregating for two hard-endosperm alleles (*PinA-D1b* and *PinB-D1b*), a significant difference in grain hardness was detected (Martin *et al.* 2001). Although small, this difference was shown to influence milling performance. Distinguishing these alleles by conventional means (single-kernel hardness tester or near-infrared spectroscopy) would be very difficult because experimental error and environmental variation could easily mask such a small effect.

7.5.2.4 Retrospective breeding and MAS

With current marker technology in wheat, it is not practical to practice MAS in the hundreds of populations handled each year by an individual breeding programme. Indeed, most reports of MAS involve few, carefully selected populations (e.g. Eagles *et al.* 2001). One possible means of expanding the use of MAS to additional populations is through what is referred to as 'retrospective breeding'. The concept is that at any given stage in a breeding programme (e.g. F_2, headrows, preliminary yield trial, etc.), the breeder can judge the genetic potential of a population or collection of lines. Those populations that are most desirable or produce the most promising lines can be resampled from remnant seed of an appropriate generation, or reconstituted entirely, and then subjected to MAS. In this way, the breeder can have some assurance that the population being subjected to the time and expense of MAS is likely to produce superior inbred lines at the end of the selection and evaluation process. This retrospective approach avoids the time and expense of conducting MAS in populations that do not produce improved lines as a result of poor combining ability or absence of genes for other required traits. The screening of additional progeny in this more directed and efficient manner, with a subsequent increase in population size, should substantially increase the possibility of recovering superior progeny and easily negate the loss of time associated with resampling or reconstructing a cross or population.

7.5.2.5 MAS conclusions

To date, MAS has been most often employed for disease resistance genes in wheat when the objective is to pyramid resistance genes or avoid difficult phenotypic screens. Also, it is probably not a coincidence that of the few examples of MAS for disease resistance in wheat, several are for genes not from the primary gene pool of *Triticum* (e.g. *Cre2*, *Lr37*, *Pch1*, *Sr38*, *Wsm1*, *Yr17*), thus increasing the likelihood of generating allele-specific primers and decreasing the likelihood of recombination between the gene and marker.

With today's technology and cost of marker application, the decision as to whether to apply MAS may be influenced more by the relative difficulty of phenotypic selection, as opposed to theoretical considerations relating to efficiency of selection. One observation that would suggest that difficulty of phenotypic selection is a major factor in a breeder's decision to implement markers are the reports of MAS in breeding for nematode resistance. As previously noted, breeding for cereal cyst nematode resistance using DNA markers is widespread in affected areas of Australia. The use of markers in breeding for resistance to the soybean cyst nematode is widespread in both public and private soybean breeding programmes (Young 1999; J. Orf, pers. commun.). The difficulty in screening for this below-ground pest and the relatively high proportion of phenotypic variance explained by single loci, are largely responsible for the prominence of this trait in MAS programmes.

Several obstacles to the implementation of MAS in wheat are gradually being overcome. There have been incremental improvements on a yearly basis to reduce cost, improve throughput and reduce the technical expertise required. SSRs are highly polymorphic PCR markers for gene mapping and selection. SNPs, preferably used in a platform that does not require PCR or a gel-based assay, may be a high-throughput means of diagnosing the presence of genes and QTL in wheat. The remaining obstacles of varying magnitude of gene effects over germplasm and environments can be addressed by:

- selecting those markers (genes) with consistent, non-epistatic effects;
- concentrating on major genes or specific resistance components; or
- transferring groups of genes.

The chances of recovering diagnostic polymorphisms can be enhanced by obtaining marker sequence from germplasm with and without the gene. As more disease resistance genes are cloned, direct allele selection (Sorrells & Wilson 1997) can be employed to search for germplasm with superior resistance alleles.

Acknowledgements

The author thanks Dr Sixin Liu for valuable comments on earlier versions of this chapter, and for providing data used in Figures 7.2, 7.4 and 7.7. Thanks are also expressed to Mike Pumphrey for providing data used in Figures 7.5 and 7.6.

References

Ahn, S., Anderson, J.A., Sorrells, M.E. & Tanksley, S.D. (1993) Homoeologous relationships of rice, wheat and maize chromosomes. *Molecular and General Genetics*, **241**, 483–490.

Altpeter, F., Diaz, I., McAuslane, H., Gaddour, K., Carbonero, P. & Vasil, I.K. (1999) Increased insect resistance in transgenic wheat stably expressing trypsin inhibitor CMe. *Molecular Breeding*, **5**, 53–63.

Alvarez, M.L., Guelman, S., Halford, N.G., *et al.* (2000) Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. *Theoretical and Applied Genetics*, **100**, 319–327.

Alvarez, M.L., Gomez, M., Carrillo, J.M. & Vallejos, R.H. (2001) Analysis of dough functionality of flours from transgenic wheat. *Molecular Breeding*, **8**, 103–108.

Anderson, J.A., Ogihara, Y., Sorrells, M.E. & Tanksley, S.D. (1992) Development of a chromosomal arm map for wheat based on RFLP markers. *Theoretical and Applied Genetics*, **83**, 1035–1043.

Anderson, J.A., Churchill, G.A., Autrique, J.E., Tanksley, S.D. & Sorrells, M.E. (1993) Optimizing parental selection for genetic linkage maps. *Genome*, **36**, 181–186.

Anderson, J.A., Stack, R.W., Liu, S., *et al.* (2001) DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, **102**, 1164–1168.

Aquino, P., Carrión, F. & Calvo, R. (1999) Selected Wheat Statistics. In: CIMMYT 1998–99 World Wheat Facts and Trends. Global Wheat Research in a Changing World: Challenges and Achievements (ed. P.L. Pingali), CIMMYT, Mexico, D.F.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Arumeganathan, K. & Earle, E.D. (1991) Estimation of nuclear DNA content of plants by flow cytometry. *Plant Molecular Biology Reports*, **9**, 229–233.

Bai, G. & Shaner, G. (1994) Scab of wheat: prospects for control. *Plant Disease*, **78**, 760–766.

Bai, G.H., Kolb, F.L., Shaner, G. & Domier, L.L. (1999) Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. *Phytopathology*, **89**, 343–348.

Barrett, B.A. & Kidwell, K.K. (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Science*, **38**, 1261–1271.

Barrett, B.A., Kidwell, K.K. & Fox, P.N. (1998) Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. *Crop Science*, **38**, 1271–1278.

Beavis, W.D. (1998) QTL analysis: power, precision and accuracy. In: *Molecular Dissection of Complex Traits* (ed. A.H. Paterson), pp. 145–161. CRC Press, Boca Raton, FL.

Becker, D., Brettschneider, R. & Lorz, H. (1994) Fertile transgenic wheat from microprojectile bombardment of scutellar tissue. *Plant Journal*, **5**, 299–307.

Blechl, A.E., Le, H.Q. & Anderson, O.D. (1998) Engineering changes in wheat flour by genetic transformation. *Journal of Plant Physiology*, **152**, 703–707.

Bliffeld, M., Mundy, J., Potrykus, I. & Futterer, J. (1999) Genetic engineering of wheat for increased resistance to powdery mildew disease. *Theoretical and Applied Genetics*, **98**, 1079–1086.

Bohn, M., Utz, H.F. & Melchinger, A.E. (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Science*, **39**, 228–237.

Bommineni, V.R., Jauhar, P.P. & Peterson, T.S. (1997) Transgenic durum wheat by microprojectile bombardment of isolated scutella. *Journal of Heredity*, **88**, 475–481.

Borner, A., Röder, M.S., Unger, O. & Meinel, A. (2000) The detection and molecular mapping of a major gene for non-specific adult-plant disease resistance against stripe rust (*Puccinia striiformis*) in wheat. *Theoretical and Applied Genetics*, **100**, 1095–1099.

Brueggeman, R., Rostoks, N., Kudrna, D., *et al.* (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences, USA*, **99**, 9328–9333.

Buerstmayr, H., Lemmens, M., Hartl, L., *et al.* (2002) Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (type II resistance). *Theoretical and Applied Genetics*, **104**, 84–91.

Burkhamer, R.L., Lanning, S.P., Martens, R.J., Martin, J.M. & Talbert, L.E. (1998) Predicting progeny variance from parental divergence in hard red spring wheat. *Crop Science*, **38**, 243–248.

Campbell, K.G., Finney, P.L., Bergman, C.J., *et al.* (2001) Quantitative trait loci associated with milling and baking quality in a soft x hard wheat cross. *Crop Science*, **41**, 1275–1285.

Chalhoub, B., Lefevre, A., Rouault, P., *et al.* (2002) Construction of a hexaploid wheat BAC library for studying genome organization and cloning of interesting genes and QTLs. In: Plant and Animal Genome X Abstracts, San Diego, CA.

Chalmers, K.J., Campbell, A.W., Kretschmer, J., *et al.* (2001) Construction of three linkage maps in bread wheat, *Triticum aestivum*. *Australian Journal of Agricultural Research*, **52**, 1089–1119.

Chantret, N., Sourdille, P., Röder, M., Tavaud, M., Bernard, M. & Doussinault, G. (2000) Location and mapping of the powdery mildew resistance gene MIRE and detection of a resistance QTL by bulked segregant analysis (BSA) with microsatellites in wheat. *Theoretical and Applied Genetics*, **100**, 1217–1224.

Chantret, N., Mingeot, D., Sourdille, P., Bernard, M., Jacquemin, J.M. & Doussinault, G. (2001) A major QTL for powdery mildew resistance is stable over time and at two development stages in winter wheat. *Theoretical and Applied Genetics*, **103**, 962–971.

Chen, W.P., Chen, P.D., Liu, D.J., *et al.* (1999) Development of wheat scab symptoms is delayed in transgenic wheat plants that constitutively express a rice thaumatin-like protein gene. *Theoretical and Applied Genetics*, **99**, 755–760.

Cheng, M., Fry, J.E., Pang, S.Z., *et al.* (1997) Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiology*, **115**, 971–980.

Clausen, M., Krauter, R., Schachermayr, G., Potrykus, I. & Sautter, C. (2000) Antifungal activity of a virally encoded gene in transgenic wheat. *Nature Biotechnology*, **18**, 446–449.

Cox, T.S. & Hatchett, J.H. (1994) Hessian fly-resistance gene H26 transferred from *Triticum tauschii* to common wheat. *Crop Science*, **34**, 958–960.

Delaney, D.E., Nasuda, S., Endo, T.R., Gill, B.S. & Hulbert, S.H. (1995a) Cytologically based physical maps of the group 3 chromosomes of wheat. *Theoretical and Applied Genetics*, **91**, 780–782.

Delaney, D.E., Nasuda, S., Endo, T.R., Gill, B.S. & Hulbert, S.H. (1995b) Cytologically based physical maps of the group-2 chromosomes of wheat. *Theoretical and Applied Genetics*, **91**, 568–573.

de la Peña, R.C. & Murray, T.D. (1994) Identifying wheat genotypes resistant to eyespot disease with a β-glucuronidase-transformed strain of *Pseudocercosporella herpotrichoides*. *Phytopathology*, **84**, 972–977.

de la Peña, R.C., Murray, T.D. & Jones, S.S. (1996) Linkage relations among eyespot resistance gene Pch2, endopeptidase Ep-A1b, and RFLP marker Xpsr121 on chromosome 7A of wheat. *Plant Breeding*, **115**, 273–275.

Delibes, A., Romero, D., Aguaded, S., *et al.* (1993) Resistance to the cereal nematode (*Heterodera avenae* Woll.) transferred from the wild grass *Aegilops ventricosa* to hexaploid wheat by a 'stepping-stone' procedure. *Theoretical and Applied Genetics*, **87**, 402–408.

Demeke, T., Laroche, A. & Gaudet, D.A. (1996) A DNA marker for the Bt-10 common bunt resistance gene in wheat. *Genome*, **39**, 51–55.

Demeke, T., Hucl, P., Baga, M., Caswell, K., Leung, N. & Chibbar, R.N. (1999) Transgene inheritance and silencing in hexaploid spring wheat. *Theoretical and Applied Genetics*, **99**, 947–953.

Devos, K.M. & Gale, M.D. (2000) Genome relationships: the grass model in current research. *Plant Cell*, **12**, 637–646.

Devos, K.M., Dubcovsky, J., Dvorak, J., Chinoy, C.N. & Gale, M.D. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theoretical and Applied Genetics*, **91**, 282–288.

Devos, K.M., Millan, T. & Gale, M.D. (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye, and barley. *Theoretical and Applied Genetics*, **85**, 784–792.

Devos, K.M., Wang, Z.M., Beales, J., Sasaki, Y. & Gale, M.D. (1998) Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theoretical and Applied Genetics*, **96**, 63–68.

Devos, K.M., Pittaway, T.S., Reynolds, A. & Gale, M.D. (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theoretical and Applied Genetics*, **100**, 190–198.

Donini, P., Law, J.R., Koebner, R.M.D., Reeves, J.C. & Cooke, R.J. (2000) Temporal trends in the diversity of UK wheat. *Theoretical and Applied Genetics*, **100**, 912–917.

Dubcovsky, J., Luo, M.-C., Zhong, G.-Y., *et al.* (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. *Genetics*, **143**, 983–999.

Dubcovsky, J., Echaide, M., Giancola, S., *et al.* (1997) Seed-storage-protein loci in RFLP maps of diploid, tetraploid, and hexaploid wheat. *Theoretical and Applied Genetics*, **95**, 1169–1180.

Dubcovsky, J., Lijavetzky, D., Appendino, L. & Tranquilli, G. (1998) Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement. *Theoretical and Applied Genetics*, **97**, 968–975.

Dubcovsky, J., Ramakrishna, W., SanMiguel, P.J., *et al.* (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiology*, **125**, 1342–1353.

Dvorak, J., DiTerlizzi, P., Zhang, H.B. & Resta, P. (1993) The evolution of polyploid wheat: identification of the A genome donor species. *Genome*, **36**, 21–31.

Eagles, H.A., Bariana, H.S., Ogbonnaya, F.C., *et al.* (2001) Implementation of markers in Australian wheat breeding. *Australian Journal of Agricultural Research*, **52**, 1349–1356.

Eastwood, R.F., Lagudah, E.S., Appels, R., Hannah, M. & Kollmorgen, J.F. (1991) Triticum tauschii: a novel source of resistance to cereal cyst nematode (*Heterodera avenae*). *Australian Journal of Agricultural Research*, **42**, 69–77.

Eastwood, R.F., Lagudah, E.S. & Appels, R. (1994) A direct search for DNA sequences tightly linked to cereal cyst nematode resistance genes in *Triticum tauschii*. *Genome*, **37**, 311–319.

Effertz, R.J., Anderson, J.A. & Francl, L.J. (2001) Restriction fragment length polymorphism mapping of resistance to two races of *Pyrenophora tritici-repentis* in adult and seedling wheat. *Phytopathology*, **91**, 572–578.

El-Assal, S.E.D., Alonso-Blanco, C., Peeters, A.J.M., Raz, V. & Koornneef, M. (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nature Genetics*, **29**, 435–440.

Endo, T.R. & Gill, B.S. (1996) The deletion stocks of common wheat. *Journal of Heredity*, **87**, 295–307.

Fabrizius, M.A., Busch, R.H., Khan, K. & Huckle, L. (1998) Genetic diversity and heterosis of spring wheat crosses. *Crop Science*, **38**, 1108–1112.

Faris, J.D., Anderson, J.A., Francl, L.J. & Jordahl, J.G. (1997) RFLP mapping of resistance to chlorosis induction by *Pyrenophora tritici-repentis* in wheat. *Theoretical and Applied Genetics*, **94**, 98–103.

Faris, J.D., Haen, K.M. & Gill, B.S. (2000) Saturation mapping of a gene-rich recombination hot spot in wheat. *Genetics*, **154**, 823–835.
Feldman, M., Lupton, F.G.H. & Miller, T.E. (1995) Wheats. In: *Evolution of Crop Plants* (eds J. Smartt & N.W. Simmonds), pp. 184–192. Longman Group, London.
Feuillet, C. & Keller, B. (1999) High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences, USA*, **96**, 8265–8270.
Feuillet, C., Schachermayr, G. & Keller, B. (1997) Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. *Plant Journal*, **11**, 45–52.
Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S. & van der Knaap, E. (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.
Gill, K.S., Gill, B.S. & Endo, T.R. (1993) A chromosome region-specific mapping strategy reveals gene-rice telomeric ends in wheat. *Chromosoma*, **102**, 374–381.
Gill, K.S., Gill, B.S., Endo, T.R. & Taylor, T. (1996a) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics*, **144**, 1883–1891.
Gill, K.S., Gill, B.S., Endo, T.R. & Boyko, E.V. (1996b) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics*, **143**, 1001–1012.
Giroux, M.J. & Morris, C.F. (1997) A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. *Theoretical and Applied Genetics*, **95**, 857–864.
Giroux, M.J, Talbert, L., Habernicht, D.K., Lanning, S., Hemphill, A. & Martin, J.M. (2000) Association of puroindoline sequence type and grain hardness in hard red spring wheat. *Crop Science*, **40**, 370–374.
Goff, S., Ricke, D., Lan, T., *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.
Gupta, P.K., Varshney, R.K., Sharma, P.C. & Amesh, B. (1999) Molecular markers and their applications in wheat breeding. *Plant Breeding*, **118**, 369–390.
Gupta, P.K., Roy, J.K. & Prasad, M. (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science*, **80**, 524–535.
Han, F., Kilian, A., Chen, J.P., *et al*. (1999) Sequence analysis of a rice BAC covering the syntenous barley Rpg1 region. *Genome*, **42**, 1071–1076.
Harker, N., Rampling, L.R., Shariflou, M.R., *et al*. (2001) Microsatellites as markers for Australian wheat improvement. *Australian Journal of Agricultural Research*, **52**, 1121–1130.
Hatchett, J.H. & Gill, B.S. (1981) D-genome sources of resistance in *Triticum tauschii* to Hessian fly *Mayetiola destructor*. *Journal of Economics and Entomology*, **72**, 126–127.
He, G.Y., Rooke, L., Steele, S., *et al*. (1999) Transformation of pasta wheat (*Triticum turgidum* L-var. *durum*) with high-molecular-weight glutenin subunit genes and modification of dough functionality. *Molecular Breeding*, **5**, 377–386.
Hohmann, U., Endo, T.R., Gill, K.S. & Gill, B.S. (1994) Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. *Molecular and General Genetics*, **245**, 644–653.
Howes, N.K., Woods, S.M. & Townley-Smith, T.F. (1998) Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs. *Euphytica*, **100**, 225–230.
Ivandic, V., Malyshev, S., Korzun, V., Graner, A. & Borner, A. (1999) Comparative mapping of a gibberellic acid-insensitive dwarfing gene (Dwf2) on chromosome 4HS in barley. *Theoretical and Applied Genetics*, **98**, 728–731.
Jiang, J., Friebe, B. & Gill, B.S. (1994) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.
Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. & Dean, C. (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science*, **290**, 344–347.

Kam-Morgan, L.N.W., Gill, B.S. & Muthukrishnan, S. (1989) DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. *Genome*, **32**, 724–732.

Kato, K., Nakamura, W., Tabiki, T., Miura, H. & Sawada, S. (2001) Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. *Theoretical and Applied Genetics*, **102**, 980–985.

Kerber, E.R. (1987) Resistance to leaf rust in hexaploid wheat: Lr32, a third gene derived from *Triticum tauschii*. *Crop Science*, **27**, 204–206.

Khan, I.A., Procunier, J.D., Humphreys, D.G., *et al*. (2000) Development of PCR-based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. *Crop Science*, **40**, 518–524.

Kojima, T. & Ogihara, Y. (1998) High-resolution RFLP map of the long arm of chromosome 5A in wheats and its synteny among cereals. *Genes & Genetic Systems*, **73**, 51–58.

Kolmer, J.A. & Dyck, P.L. (1994) Gene expression in the *Triticum aestivum-Puccinia recondita* f. sp. *tritici* gene-for-gene system. *Phytopathology*, **84**, 437–440.

Kota, R.S., Gill, K.S. & Gill, B.S. (1993) A cytogenetically based physical map of chromosome 1B in common wheat. *Genome*, **36**, 548–554.

Lagudah, E.S., Moullet, O. & Appels, R. (1997) Map-based cloning of a gene sequence encoding a nucleotide binding domain and a leucine-rich region at the Cre3 nematode resistance locus of wheat. *Genome*, **40**, 659–665.

Lagudah, E.S., Moullet, O., Ogbonnaya, F., *et al*. (1998) Cloning of disease resistance gene sequences at loci conferring cyst nematode resistance genes in wheat. In: Proceedings, 9th International Wheat Genetics Symposium, Saskatoon, Vol. 1 (ed. A.E. Slinkard), pp. 184–186. University Extension Press, Saskatoon, Canada.

Langridge, P. & Chalmers, K. (1998) Techniques for marker development. In: Proceedings, 9th International Wheat Genetics Symposium, Saskatoon, Vol. 1 (ed. A.E. Slinkard), pp. 107–117. University Extension Press, Saskatoon, Canada.

Langridge, P., Lagudah, E.S., Holton, T.A., Appels, R., Sharp, P.J. & Chalmers, K.J. (2001) Trends in genetic and genome analyses in wheat: a review. *Australian Journal of Agricultural Research*, **52**, 1043–1077.

Lee, Y.K., Bekes, F., Gupta, R., Appels, R. & Morell, M.K. (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. I. Variation in A-genome species. *Theoretical and Applied Genetics*, **98**, 119–125.

Lelley, T., Stachel, M., Grausgruber, H. & Vollmann, J. (2000) Analysis of relationships between *Aegilops tauschii* and the D genome of wheat utilizing microsatellites. *Genome*, **43**, 661–668.

Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R. & Dubcovsky, J. (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. *Genome*, **42**, 1176–1182.

Liu, S. & Anderson, J.A. (2003) Marker assisted evaluation of *Fusarium* head blight resistant wheat germplasm. *Crop Science*, **43**, 760–766.

Liu, Y.G., Nagaki, K., Fujita, M., Kawaura, K., Uozumi, M. & Ogihara, Y. (2000) Development of an efficient maintenance and screening system for large-insert genomic DNA libraries of hexaploid wheat in a transformation-competent artificial chromosome (TAC) vector. *Plant Journal*, **23**, 687–695.

Liu, Z.Q., Pei, Y. & Pu, Z.J. (1999) Relationship between hybrid performance and genetic diversity based on RAPD markers in wheat, *Triticum aestivum* L. *Plant Breeding*, **118**, 119–123.

Lookhart, G.L., Hagman, K. & Kasarda, D.D. (1993) High-molecular-weight glutenin subunits of the most commonly grown wheat cultivars in the U.S. in 1984. *Plant Breeding*, **110**, 48–62.

Lorz, H., Becker, D. & Lutticke, S. (1998) Molecular wheat breeding by direct gene transfer. *Euphytica*, **100**, 219–223.

Lotti, C., Salvi, S., Pasqualone, A., Tuberosa, R. & Blanco, A. (2000) Integration of AFLP markers into an RFLP-based map of durum wheat. *Plant Breeding*, **119**, 393–401.

Loughman, R., Lagudah, E.S., Trottet, M., Wilson, R.E. & Mathews, A. (2001) *Septoria nodorum* blotch resistance in *Aegilops tauschii* and its expression in synthetic amphiploids. *Australian Journal of Agricultural Research*, **52**, 1393–1402.

Marino, C.L., Nelson, J.C., Lu, Y.H., *et al.* (1996) Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L em Thell). *Genome*, **39**, 359–366.

Martin, J.M., Frohberg, R.C., Morris, C.F., Talbert, L.E. & Giroux, M.J. (2001) Milling and bread baking traits associated with puroindoline sequence type in hard red spring wheat. *Crop Science*, **41**, 228–234.

Martins-Lopes, P., Zhang, H. & Koebner, R. (2001) Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels. *Plant Molecular Biology Reports*, **19**, 159–162.

McFadden, E.S. & Sears, E.R. (1946) The origin of *Triticum speltoides* and its free-threshing hexaploid relatives. *Journal of Heredity*, **37**, 107–116.

McIntosh, R.A., Hart, G.E., Devos, K.M., Gale, M.D. & Rogers, W.J. (1998) Catalogue of gene symbols for wheat. In: Proceedings, 9th International Wheat Genetics Symposium, Saskatoon, Vol. 5 (ed. A.E. Slinkard), pp. 1–235. University Extension Press, Saskatoon, Canada.

McKendry, A.L. & Henke, G.E. (1994) Evaluation of wheat wild relatives for resistance to *Septoria tritici* blotch. *Crop Science*, **34**, 1080–1084.

McMillin, D.E., Allan, R.E. & Roberts, D.E. (1986) Association of an isozyme locus and strawbreaker foot rot resistance derived from *Aegilops ventricosa* in wheat. *Theoretical and Applied Genetics*, **72**, 743–747.

Mercado, L.A., Souza, E. & Kephart, K.D. (1996) Origin and diversity of North American hard spring wheats. *Theoretical and Applied Genetics*, **93**, 593–599.

Mesfin, A., Frohberg, R.C. & Anderson, J.A. (1999) RFLP markers associated with high grain protein from *Triticum turgidum* L. var. *dicoccoides* introgressed into hard red spring wheat. *Crop Science*, **39**, 508–513.

Metakovsky, E.V. & Branlard, G. (1998) Genetic diversity of French common wheat germplasm based on gliadin alleles. *Theoretical and Applied Genetics*, **96**, 209–218.

Metakovsky, E.V., Gomez, M., Vazquez, J.F. & Carrillo, J.M. (2000) High genetic diversity of Spanish common wheats as judged from gliadin alleles. *Plant Breeding*, **119**, 37–42.

Mickelsonyoung, L., Endo, T.R. & Gill, B.S. (1995) A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. *Theoretical and Applied Genetics*, **90**, 1007–1011.

Moore, G., Devos, K.M., Wang, Z. & Gale, M.D. (1995) Grasses, line up and form a circle. *Current Biology*, **5**, 737–739.

Morris, C.F., Lillemo, M., Simeone, M.C., Giroux, M.J., Babb, S.L. & Kidwell, K.K. (2001) Prevalence of puroindoline grain hardness genotypes among historically significant North American spring and winter wheats. *Crop Science*, **41**, 218–228.

Moullet, O., Zhang, H.B. & Lagudah, E.S. (1999) Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theoretical and Applied Genetics*, **99**, 305–313.

Nagaoka, T. & Ogihara, Y. (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theoretical and Applied Genetics*, **94**, 597–602.

Nakamura, H. (2000) Allelic variation at high-molecular-weight glutenin subunit loci, Glu-A1, Glu-B1and Glu-D1, in Japanese and Chinese hexaploid wheats. *Euphytica*, **112**, 187–193.

Naranjo, T., Roca, A., Goicoechea, P.G. & Giraldez, R. (1987) Arm homoeology of wheat and rye chromosomes. *Genome*, **29**, 873–882.

Nehra, N.S., Chibbar, R.N., Leung, N., *et al.* (1994) Self-fertile transgenic wheat plants regenerated from isolated scutellar tissues following microprojectile bombardment with two distinct gene constructs. *Plant Journal*, **5**, 285–297.

Nelson, J.C., Sorrells, M.E., Van Deynze, A.E., *et al.* (1995a) Molecular mapping of wheat – major genes and rearrangements in homoeologous groups 4, 5, and 7. *Genetics*, **141**, 721–731.

Nelson, J.C., Van Deynze, A.E., Autrique, E., *et al.* (1995b) Molecular mapping of wheat – homoeologous group 2. *Genome*, **38**, 516–524.

Nelson, J.C., Van Deynze, A.E., Autrique, E., *et al.* (1995c) Molecular mapping of wheat – homoeologous group 3. *Genome*, **38**, 525–533.

Nelson, J.C., Singh, R.P., Autrique, J.E. & Sorrells, M.E. (1997) Mapping genes conferring and suppressing leaf rust resistance in wheat. *Crop Science*, **37**, 1928–1935.

Nelson, J.C., Autrique, J.E., Fuentes-Dávila, G. & Sorrells, M.E. (1998) Chromosomal location of genes for resistance to karnal bunt in wheat. *Crop Science*, **38**, 231–236.

Ogbonnaya, F.C., Seah, S., Delibes, A., *et al.* (2001a) Molecular-genetic characterization of a new nematode resistance gene in wheat. *Theoretical and Applied Genetics*, **102**, 623–629.

Ogbonnaya, F.C., Subrahmanyam, N.C., Moullet, O., *et al.* (2001b) Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat. *Australian Journal of Agricultural Research*, **52**, 1367–1374.

Oldach, K.H., Becker, D. & Lorz, H. (2001) Heterologous expression of genes mediating enhanced fungal resistance in transgenic wheat. *Molecular Plant-Microbe Interactions*, **14**, 832–838.

Osborne, B.G., Turnbull, K.M., Anderssen, R.S., Rahman, S., Sharp, P.J. & Appels, R. (2001) The hardness locus in Australian wheat lines. *Australian Journal of Agricultural Research*, **52**, 1275–1286.

Pagnotta, M.A., Nevo, E., Beiles, A. & Porceddu, E. (1995) Wheat storage proteins – glutenin diversity in wild emmer, *Triticum dicoccoides*, in Israel and Turkey. 2. DNA diversity detected by PCR. *Theoretical and Applied Genetics*, **91**, 409–414.

Payne, P.I., Nightingale, M.A., Krattiger, A.F. & Holt, L.M. (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *Journal of the Science of Food and Agriculture*, **40**, 51–65.

Peng, J.R., Carol, P., Richards, D.E., *et al.* (1997) The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes and Development*, **11**, 3194–3205.

Peng, J.R., Richards, D.E., Hartley, N.M., *et al.* (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400** 256–261.

Peng, J., Korol, A.B., Fahima, T., *et al.* (2000a) Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. *Genome Research*, **10**, 1509–1531.

Peng, J.H., Fahima, T., Röder, M.S., *et al.* (2000b) High-density molecular map of chromosome region harboring stripe-rust resistance genes YrH52 and Yr15 derived from wild emmer wheat, *Triticum dicoccoides. Genetica*, **109**, 199–210.

Peng, J.H., Fahima, T., Röder, M.S., Li, Y.C., Grama, A. & Nevo, E. (2000c) Microsatellite high-density mapping of the stripe rust resistance gene YrH52 region on chromosome 1B and evaluation of its marker-assisted selection in the F-2 generation in wild emmer wheat. *New Phytologist*, **146**, 141–154.

Perenzin, M., Corbellini, M., Accerbi, M., Vaccino, P. & Borghi, B. (1998) Bread wheat – F1 hybrid performance and parental diversity estimates using molecular markers. *Euphytica*, **100**, 273–279.

Pestsova, E., Ganal, M.W. & Röder, M.S. (2000a) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. *Genome*, **43**, 689–697.

Pestsova, E., Korzun, V., Goncharov, N.P., Hammer, K., Ganal, M.W. & Röder, M.S. (2000b) Microsatellite analysis of *Aegilops tauschii* germplasm. *Theoretical and Applied Genetics*, **101**, 100–106.

Pfluger, L.A., Martin, L.M. & Alvarez, J.B. (2001) Variation in the HMW and LMW glutenin subunits from Spanish accessions of emmer wheat (*Triticum turgidum* ssp. *dicoccum* Schrank). *Theoretical and Applied Genetics*, **102**, 767–772.

Plaschke, J., Ganal, M.W. & Röder, M.S. (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theoretical and Applied Genetics*, **91**, 1001–1007.

Popineau, Y., Deshayes, G., Lefebvre, J., Fido, R., Tatham, A.S. & Shewry, P.R. (2001) Prolamin aggregation, gluten viscoelasticity, and mixing properties of transgenic wheat lines expressing 1Ax and 1Dx high molecular weight glutenin subunit transgenes. *Journal of Agriculture and Food Chemistry*, **49**, 395–401.

Pretorius, Z.A. & Roelfs, A.P. (1996) The role of Lr10, Lr13, and Lr34 in the expression of adult-plant resistance to leaf rust in the wheat cultivar Era. *Plant Disease*, **80**, 199–202.

Rafalski, A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, **5**, 94–100.

Rajaram, S. (1999) Approaches for breaching yield stagnation in wheat. *Genome*, **42**, 629–634.

Raupp, W.J., Sukhwinder, S., Brown-Guedira, G.L. & Gill, B.S. (2001) Cytogenetic and molecular mapping of the leaf rust resistance gene Lr39 in wheat. *Theoretical and Applied Genetics*, **102**, 347–352.

Riede, C.R. & Anderson, J.A. (1996) Linkage of RFLP markers to an aluminum tolerance gene in wheat. *Crop Science*, **36**, 905–909.

Röder, M.S., Korzun, V., Gill, B.S. & Ganal, M.W. (1998a) The physical mapping of microsatellite markers in wheat. *Genome*, **41**, 278–283.

Röder, M.S., Korzun, V., Wendehake, K., *et al.* (1998b) A microsatellite map of wheat. *Genetics*, **149**, 2007–2023.

Rodriguezquijano, M., Nietotaladriz, M.T. & Carrillo, J.M. (1997) Variation in b-LMW glutenin subunits in einkorn wheats. *Genetic Resources and Crop Evolution*, **44**, 539–543.

Rooke, L., Bekes, F., Fido, R., *et al.* (1999) Overexpression of a gluten protein in transgenic wheat results in greatly increased dough strength. *Journal of Cereal Science*, **30**, 115–120.

Sandhu, D. & Gill, K.S. (2002) Gene-containing regions of wheat and the other grass genomes. *Plant Physiology*, **128**, 803–811.

Sandhu, D., Champoux, J.A., Bondareva, S.N. & Gill, K.S. (2001) Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. *Genetics*, **157**, 1735–1747.

Sarma, R.N., Fish, L., Gill, B.S. & Snape, J.W. (2000) Physical characterization of the homoeologous group 5 chromosomes of wheat in terms of rice linkage blocks, and physical mapping of some important genes. *Genome*, **43**, 191–198.

Sasaki, T. (2001) Rice genome analysis to understand the rice plant as an assembly of genetic codes. *Photosynthesis Research*, **70**, 119–127.

Sears, E.R. (1954) The aneuploids of common wheat. *Missouri Agricultural Experiment Station Research Bulletin*, **572**, 1–58.

Sears, E.R. (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: *Chromosome Manipulation and Plant Genetics* (eds R. Riley & K.R. Lewis), pp. 29–45. Oliver and Boyd, Edinburgh.

Sears, E.R. & Sears, L.M.S. (1979) The telocentric chromosomes of common wheat. In: *Proceedings, 5th International Wheat Genetics Symposium* (ed. S. Ramanujam), pp. 389–407. Indian Society of Genetics and Plant Breeding, New Delhi.

Shewry, P.R., Tatham, A.S., Fido, R., Jones, H., Barcelo, P. & Lazzeri, P.A. (2001) Improving the end use properties of wheat by manipulating the grain protein composition. *Euphytica*, **119**, 45–48.

Singh, R.P., Nelson, J.C. & Sorrells, M.E. (2000) Mapping Yr28 and other genes for resistance to stripe rust in wheat. *Crop Science*, **40**, 1148–1155.

Sivamani, E., Brey, C.W., Dyer, W.E., Talbert, L.E. & Qu, R.D. (2000) Resistance to wheat streak mosaic virus in transgenic wheat expressing the viral replicase (NIb) gene. *Molecular Breeding*, **6**, 469–477.

Smith, D.B. & Flavell, R.B. (1975) Characterisation of the wheat genome by renaturation kinetics. *Chromosoma*, **50**, 223–242.

Sorrells, M.E. & Wilson, W.A. (1997) Direct classification and selection of superior alleles for crop improvement. *Crop Science*, **37**, 691–697.

Spielmeyer, W., Huang, L., Bariana, H., Laroche, A., Gill, B.S. & Lagudah, E.S. (2000a) NBS-LRR sequence family is associated with leaf and stripe rust resistance on the end of homoeologous chromosome group 1S of wheat. *Theoretical and Applied Genetics*, **101**, 1139–1144.

Spielmeyer, W., Moullet, O., Laroche, A. & Lagudah, E.S. (2000b) Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. *Genetics*, **155**, 361–367.

Stein, N., Feuillet, C., Wicker, T., Schlagenhauf, E. & Keller, B. (2000) Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the Lr10 resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proceedings of the National Academy of Sciences, USA*, **97**, 13436–13441.

Stoger, E., Williams, S., Christou, P., Down, R.E. & Gatehouse, J.A. (1999) Expression of the insecticidal lectin from snowdrop (*Galanthus nivalis* agglutinin; GNA) in transgenic wheat plants: effects on predation by the grain aphid *Sitobion avenae*. *Molecular Breeding*, **5**, 65–73.

Symes, K.J. (1965) The inheritance of grain hardness in wheat as measured by the particle size index. *Australian Journal of Agricultural Research*, **16**, 113–123.

Takahashi, Y., Shomura, A., Sasaki, T. & Yano, M. (2001) Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences, USA*, **98**, 7922–7927.

Talbert, L.E., Smith, L.Y. & Blake, M.K. (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome*, **41**, 402–407.

Tang, Y., Sorrells, M.E., Kochian, L.V. & Garvin, D.F. (2000) Identification of RFLP markers linked to the barley aluminum tolerance gene Alp. *Crop Science*, **40**, 778–782.

van Beuningen, L.T. & Busch, R.H. (1997) Genetic diversity among North American spring wheat cultivars. 2. Ancestor contributions to gene pools of different eras and regions. *Crop Science*, **37**, 580–585.

Van Deynze, A.E., Dubcovsky, J., Gill, K.S., *et al.* (1995a) Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. *Genome*, **38**, 45–59.

Van Deynze, A.E., Nelson, J.C., Yglesias, E.S., *et al.* (1995b) Comparative mapping in grasses – wheat relationships. *Molecular and General Genetics*, **248**, 744–754.

Van Deynze, A.E., Sorrells, M.E., Park, W.D., *et al.* (1998) Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics*, **97**, 356–369.

Van Sanford, D., Anderson, J., Campbell, K., *et al.* (2001) Discovery and deployment of molecular markers linked to *Fusarium* head blight resistance: an integrated system for wheat and barley. *Crop Science*, **41**, 638–644.

Vasil, I.K. & Anderson, O.D. (1997) Genetic engineering of wheat gluten. *Trends in Plant Science*, **2**, 292–297.

Vasil, I.K., Bean, S., Zhao, J.M., *et al.* (2001) Evaluation of baking properties and gluten protein composition of field grown transgenic wheat lines expressing high molecular weight glutenin gene 1Ax1. *Journal of Plant Physiology*, **158**, 521–528.

Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W. & Frohberg, R.C. (1999) RFLP mapping of QTL for *Fusarium* head blight resistance in wheat. *Crop Science*, **39**, 805–811.

Wang, G., Snape, J.W., Hu, H. & Rogers, W.J. (1993) The high-molecular-weight glutenin subunit compositions of Chinese bread wheat varieties and their relationship with bread-making quality. *Euphytica*, **68**, 205–212.

Weeks, J.T., Anderson, O.D. & Blechl, A.E. (1993) Rapid production of multiple independent lines of fertile transgenic wheat (*Triticum aestivum*). *Plant Physiology*, **102**, 1077–1084.

Weng, Y., Tuleen, N.A. & Hart, G.E. (2000) Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L-em Thell.). *Theoretical and Applied Genetics*, **100**, 519–527.

Werner, J.E., Endo, T.R. & Gill, B.S. (1992) Toward a cytogenetically based physical map of the wheat genome. *Proceedings of the National Academy of Sciences, USA*, **89**, 11307–11311.

William, H.M., Hoisington, D., Singh, R.P. & González-de-León, D. (1997) Detection of quantitative trait loci associated with leaf rust resistance in bread wheat. *Genome*, **40**, 253–260.

Williams, K.J., Fisher, J.M. & Langridge, P. (1994) Identification of RFLP markers linked to the cereal cyst nematode resistance gene (Cre) in wheat. *Theoretical and Applied Genetics*, **89**, 927–930.

Williams, K.J., Fisher, J.M. & Langridge, P. (1996) Development of a PCR-based allele-specific assay from an RFLP probe linked to resistance to cereal cyst nematode in wheat. *Genome*, **39**, 798–801.

Worland, A.J., Law, C.N., Hollins, T.W., Koebner, R.M.D. & Giura, A. (1988) Location of a gene for resistance to eyespot (*Pseudocercosporella herpotrichoides*) on chromosome 7D of bread wheat. *Plant Breeding*, **101**, 43–51.

Yano, M., Katayose, Y., Ashikari, M., *et al.* (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell*, **12**, 2473–2483.

Young, N.D. (1999) A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding*, **5**, 505–510.

Yu, J., Hu, S., Wang, J., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.

Yuan, Q.P., Quackenbush, J., Sultana, R., Pertea, M., Salzberg, S.L. & Buell, C.R. (2001) Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiology*, **125**, 1166–1174.

Zhang, L.Y., French, R., Langenberg, W.G. & Mitra, A. (2001) Accumulation of barley stripe mosaic virus is significantly reduced in transgenic wheat plants expressing a bacterial ribonuclease. *Transgenic Research*, **10**, 13–19.

Zhou, H., Arrowsmith, J.W., Fromm, M.E., *et al.* (1995) Glyphosate-tolerant Cp4 and Gox genes as a selectable marker in wheat transformation. *Plant Cell Reports*, **15**, 159–163.

8 Genomics and molecular breeding for root and tuber crop improvement

M.W. Bonierbale, R. Simon, D.P. Zhang, M. Ghislain, C. Mba and X.-Q. Li

8.1 Root and tuber crop profile

Root and tuber (R&T) crops comprise a botanically diverse group of starchy staples that are critically important for food and agriculture. More than 30 species of R&T are grown in the world today, several of which play increasing roles in income generation. Foremost among these in terms of aggregate output and estimated value of production are potato (*Solanum tuberosum* L.), cassava (*Manihot esculenta* Crantz), sweetpotato [*Ipomoea batatas* L. (Lam)] and yam (*Dioscorea* spp.). Potato and cassava are from the Andean centre of origin of cultivated plants, and sweetpotato is from the Central American centre (Vavilov 1992). Yam includes several species that originated in South-east Asia, West Africa and tropical America. Some species have moved from Africa to North and South America, and others have travelled from Asia to Africa (Hahn *et al.* 1987; Asiedu *et al.* 1997). Other prominent R&T crops include cocoyam, taro and yam bean, as well as additional Andean species such as arracacha, mashua, oca and ulluco. Nearly all are complex polyploids and are vegetatively propagated.

The four major R&T crops – potato, cassava, sweetpotato and yam – occupy about 50 million hectares worldwide. Farmers produce about 640 million metric tons of these crops annually, some 70% of which are harvested in developing countries. The production of the four major R&T crops in developing countries alone had an estimated annual value of more than US$41 billion between 1995 and 1997, nearly one-fourth the value of the major cereals (Scott *et al.* 2000).

Many of the developing world's poorest producers and most undernourished households depend on R&T crops as a contributory (if not principal) source of food and nutrition. In part, developing country production of R&T crops is increasing because they provide the highest food energy production rates per unit area of land, water and time and offer stable yields under extreme conditions under which other crops may fail (Alexandratos 1995) (Figures 8.1 and 8.2). The major part of world production is obtained with few or no inputs, under traditional cropping systems. Cassava in particular is recognized as tolerant of acid soils (Howeler 1991) and long dry seasons (El-Sharkawy 1993). In addition, some R&T crops are an important source of vitamins, essential amino acids such as lysine, and minerals (K, P, Fe, Mg and Ca) (Woolfe 1992; Spencer and Associates 1997; Kochhar 1998). R&T crops also provide an important source of employment and income in rural, often marginal areas, including for women because of their adaptation to a wide range of needs and

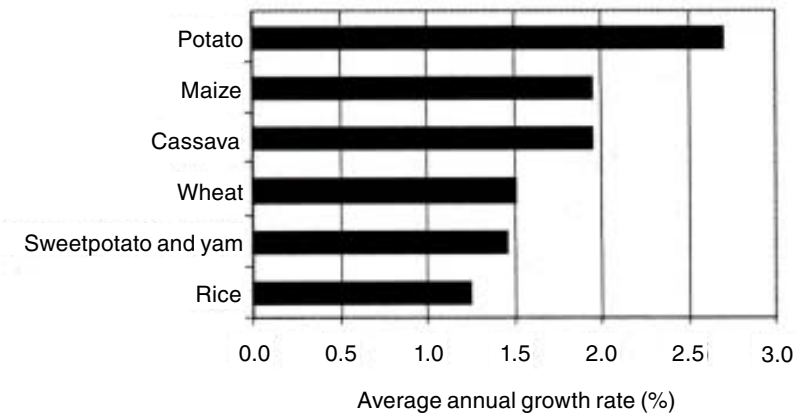


Figure 8.1 Projected growth rates for major food crops in developing countries, 1993 to 2020.

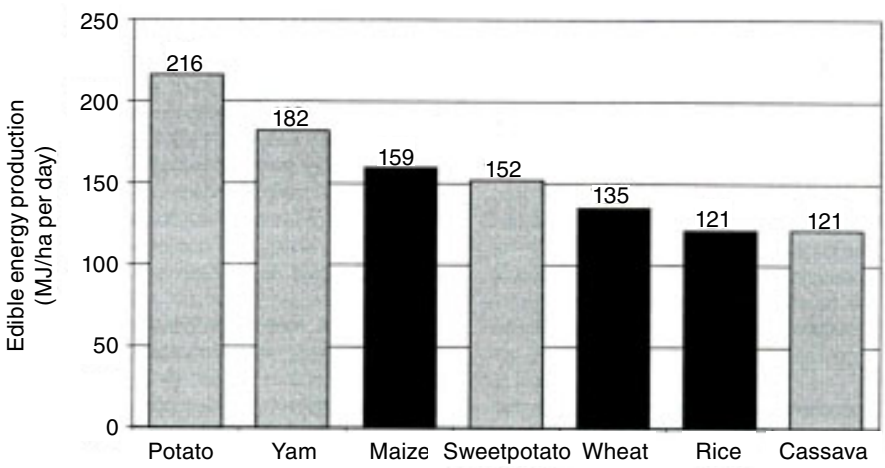


Figure 8.2 Production of edible energy from roots, tubers and major cereal crops.

market opportunities, from food security crop to cash crops, from food to feed crops, from the latter to raw materials for industrial use, and from fresh food to high-end processed products (Scott *et al*. 2000).

Root and tuber crops receive little consideration from private breeding interests, although some attention is given to potato by multinational processors and private breeders in Europe. Recently, a number of private foundations have recognized the critical role of R&T crops in development and have begun to contribute to global research agendas.

National and international collections of R&T germplasm comprise a wealth of genetic resources that are key to improving production in marginal conditions.

Population improvement is generally used to incorporate genetic diversity into advanced germplasm of R&T crops. The objectives of population improvement are to shift population means for quantitative traits in the desired direction and increase frequencies of desirable alleles to increment probabilities for the selection of successful varieties. While attention is paid to pedigree to avoid inbreeding, parental clones are often discarded and this impedes the accurate tracing of lineages.

The major focus of the germplasm conservation and crop improvement programmes in international centres is germplasm enhancement. Superior progenitors are often the more valuable outputs of international programmes than are finished varieties, because of their ability to channel new diversity and needed traits into local breeding pools. Only a small fraction of advanced breeding lines or varieties is qualified to be released as promising progenitors because of the requirement for good general combining ability. Tests of combining ability are often carried out in parallel with the main selection programme (Tai & Young 1984; Mihovilovich *et al.* 2001).

8.2 The power of comparative mapping for R&T crops

Medium- to high-density molecular genetic maps are now available for most important crops, including cassava (Fregene *et al.* 1997), sweetpotato (Kriegner *et al.* 2001), potato (Bonierbale *et al.* 1988; Rouppe van der Voort *et al.* 1998) and yam (Mignouna *et al.* 2002). Several molecular genetic maps have been published for potato, and efforts to use common anchor markers have been made to ensure that information generated in separate populations, including the highly syntenic tomato genome (Tanksley *et al.* 1992), is compatible. Use of the same population to map different 'positive' and 'negative' traits has facilitated understanding of the linkage drag often met when using wild genetic resources (e.g. Plaisted *et al.* 1993) by conventional breeding, and of the relationship among biochemical, morphological and other less-understood components of complex traits such as partial resistance (Yencho *et al.* 2000). A total of 19 genes for resistance to pathogens and pests has been localized on a range of different potato maps, several of them residing in so-called 'hot spots' for resistance (Grube *et al.* 2000; Gebhardt & Valkonen 2001). Bridge and anchor markers have also been used to unite cassava maps based on different genetic resources, permitting the localization of genes and quantitative trait loci (QTL) for resistance to the crop's most important pathogens such as cassava mosaic disease and bacterial blight (Jorge *et al.* 2000; Akano *et al.* 2002).

However, overall progress of linkage and QTL mapping in R&T crops has been slow compared with other major crop species. This is partially due to heterozygosity and polyploidy. Algorithms for linkage analysis based on the estimation of recombination in diploids (e.g. Lander *et al.* 1987) are only applicable to a limited extent in polyploids, which may undergo both bivalent and multivalent pairing at meiosis. Pseudo-testcross models have been used to monitor single-dose polymorphisms in polyploid mapping populations (Wu *et al.* 1992; Wang *et al.* 1998). However, for quantitative trait mapping it is difficult to quantify the effect of an individual allelic

substitution in a segregating heterozygous background (Conner *et al.* 1998). Additionally, the identification of important QTL alleles is more difficult in polyploids because of assumption that they would have the same single-dose segregation pattern as linked markers (Wang *et al.* 1998). Double reduction (two sister chromatids of a chromosome sort into the same gamete) due to multivalent pairings in meiosis (Fisher 1947) adds another complexity to QTL mapping in polyploids, and was invoked as the best model to explain resistance to yam mosaic virus mapped in the allotetraploid Guinea yam (*Dioscorea rotundada* Poir.) by Mignouna *et al.* (2002). The construction of several potato maps at reduced ploidy level (2× instead of 4×) has circumvented some of these difficulties, but adequate stocks are not available for diploid mapping in cassava, sweetpotato or yam. Furthermore, where it has been investigated, the consistency of marker-trait associations across ploidy levels has not been entirely satisfactory (C. Yencho, unpublished data), perhaps due to intra- and intergenic epistatic interactions.

Comparative genetic mapping of distantly related species in several plant families has revealed the orthologous nature of genes controlling similar phenotypes in different crops, relative recombination rates between parents and among species, and patterns of duplication and domestication (Ahn & Tanksley 1993; Paterson *et al.* 1995; Bernacchi & Tanksley 1997). In any crop group, the collation of genetic information across species considerably enhances studies of any one member (Van Deynze *et al.* 1998). Certainly, comparative genetics – across species of the same family or on a basis of shared traits – offers to advance research in important but under-researched species such as the R&T crops.

However, while modern studies using molecular marker systems have demonstrated colinearity and the relative conservation of gene function, it is also known that many factors contribute to chromosomal rearrangements (Paterson *et al.* 2000) which may affect gene regulation. This tends to weigh against the usefulness of cross-species information, the more so the more distant the compared species are. Yet results from human genomics (based on studies of the defence-related gene complex, MHC) suggest that some sets of genes tend to be in the same cluster across species whereas others do not: a plausible cause could be that clustered genes gain advantages by also being functional related (co-regulated), or at least having similar transcription requirements (co-transcribed) (Trowsdale 2002).

Comparisons of the *Arabidopsis* genome with sequences from other flowering plants have revealed that substantial colinearity exists between species in the arrangement of genes within chromosomal blocks (Ku *et al.* 2000). This colinearity (in short sequences at the Megabase scale) should be of value for transfer of information from model organisms whose complete genome sequences are known to identify candidate genes in defined chromosomal regions of R&T crops. Already, the extensive genetic homology among systems of host–pathogen interaction has provided short cuts to the identification of resistance genes in the Solanaceae and other species (Kanazin *et al.* 1996; Leister *et al.* 1996). Through comparative genetics combined with sequence information, markers located on individual species maps can be aligned with those of other taxa, associated with putative gene functions, and

eventually provide the basis for comparisons of allelic divergence across genera (Van Deynze *et al.* 1998). Expression studies like microarray-based results will be highly complementary to such information in order to identify co-regulated gene clusters that are probably also positionally related. In the Solanaceae, exploratory use of potato microarrays based on expressed sequence tags (ESTs) has shown promise for gene discovery in related species (Smart *et al.* 2002; B. Baker, pers. commun.), suggesting that genomic tools may have application across species that share critical biochemical or metabolic pathways. In cases of cross-species expression studies aimed at important agronomic traits, it would seem advisable to establish parallel expression experiments for both the reference and the target species.

8.3 The '…ics' technology to unravel R&T gene networks

Current efforts in genomics for R&T crops are dedicated to developing the first sets of EST libraries to facilitate germplasm evaluation and breeding. ESTs form the basis for obtaining a global 'snapshot' of gene expression (level and complexity) in a given tissue at a given time under given conditions, and therefore represent the status of the activities of enzymes encoding for specific plant metabolic pathways. This strategy of partially sequencing randomly selected cDNA clones has since evolved into an inexpensive and efficient gene discovery methodology (Ohlrogge & Benning 2000). As far back as 1992, Uchimiya *et al.* had successfully matched the sequences obtained from a rice cDNA library to reported genes in the Genebank. It is little wonder then that EST databases are the fastest growing, and constitute the largest portion of public DNA sequence databases, with approximately 1.3 million ESTs spread over most of the plant genera (Walbot & Delseny 2002). The large-scale exploitation of ESTs as reservoirs for gene cloning, evaluation of tissue-specific gene expression, as markers for map-based cloning and for the annotation genomic sequences has not only become routine in *Arabidopsis* (Györgyey *et al.* 2000; White *et al.* 2000) but has also been extended to other plant species such as poplar (Sterky *et al.* 1998), grapes (Ablett *et al.* 2000) and *Pinus* (Cato *et al.* 2001). Linked with such a novel high-throughput platform as the DNA microarray technology, ESTs become even more potent as a rapid route to the identification of genes and to linking sequence information to biological function (Richmond & Somerville 2000).

In the cassava EST project, CIAT and the French Universities of Perpignan and Montpellier have initiated collaboration aimed at the identification of starch genes. It is intended that these ESTs will be located on framework molecular and QTL maps, thereby providing the basis for linking sequence information with function. The cassava starch EST project aims at the development of at least 5000 ESTs each from two root cDNA libraries, sourced respectively from the high- and low-starch varieties CM 523-7 and MPer 183. It is hoped that the annotation of these ESTs, in concert with microarray technology, will accomplish the assignment of function to the sequences. This will ultimately lead to better exploitation of the genetic diversity available in cassava, such as expansion of the range of potential uses of cassava

starch. The scope of this start-up project is limited at present, and a larger project based on cDNAs from a broader range of cassava varieties in terms of starch quantity and quality yields will be needed to accomplish practical objectives. The resulting ESTs will be used to develop DNA chips, and converted to PCR-based markers for use in saturating and converting existing molecular genetic framework maps to functional maps, as has been demonstrated for *Pinus* (Cato *et al.* 2001) and also as markers for use in germplasm fingerprinting.

Genomic research in sweetpotato has also begun recently, but is piecemeal, as there is as yet no community-based effort for resource development. A *Bam*HI BAC library with an average insert size of 120 kbp has been constructed at North Carolina State University (He *et al.* 2001). cDNA libraries using mRNAs from specific stages in sweetpotato root development are being developed to focus on rapid bulking and starch accumulation. The first two such libraries have been constructed in a joint project between Centro Internacional de la Papa (CIP) and the Potato Research Center of Agriculture and Agri-Food Canada. So far, about 1000 ESTs have been developed (Li *et al.* 2001). A parallel project between CIP and the Austrian Research Center, Seibersdorf is oriented to EST and single nucleotide polymorphism (SNP) development. The mid-term goal of these projects is to discover genes that are involved in starch accumulation, storage root development and other priority traits.

Efforts to develop and apply genomic approaches in the Solanaceae have recently been concerted through a collaborative potato project funded by the US National Science Foundation (NSF), building on longer-standing projects on tomato and other crops. Two genome projects in the US (NSF: tomato, potato), a new Canadian Potato Genome Project and significant progress in European institutes [e.g. Scottish Crop Research Institute (SCRI), Max Planck Institute (MPI) and John Innes Centre (JIC)], following long-term classical interest in molecular biochemistry of starch, wound healing, etc., and the availability of saturated syntenic maps in the Solanaceae since the late 1980s (Bonierbale *et al.* 1988; Gebhardt *et al.* 1991; Tanksley *et al.* 1992; Prince *et al.* 1993) poise the potato to benefit exponentially from genomic tools in the next few years. In June 2002, the potato ranked 23rd among plant species in the number of EST sequences contained in dbEST, at 79 199 (TIGR/StGI (*Solanum tuberosum* gene index): 78934), while some 148 338 sequences were available for tomato. More than 30 000 potato ESTs have been contributed since the beginning of 2002. Eight core libraries (each >4000 sequences) contribute 95% of the ESTs. Microarrays of 10 000 clones are available from The Institute for Genome Research (TIGR) following construction of the StGI. Inclusion of potato ESTs in the TIGR Orthologous Gene Alignment Database (TOGA, now renamed EGO: Eucaryotic Gene Index) has permitted the identification of candidate orthologues among 13 plant gene indexes, and the in-silico anchoring of >2500 potato sequences to the *Arabidopsis* map, providing a new resource for applications of synteny (NSF potato genome project 2001). Additional core libraries for tuber development and post-harvest quality are proposed by the Canadian Potato Genome Project, as are complementary approaches to maximize gene discovery, including the construction of an activation-tagged mutant library and identification of tuber genes by expression

analysis (B. Flinn, pers. commun.). The main challenge ahead remains to identify the function of the numerous EST and gene sequences identified. Efficient in-silico identification through homology searches will still have to be complemented as well as corroborated by high-throughput reverse genetics such as transposon tagging or gene silencing (Michelmore 2000).

8.4 Gene transfer

Breeders working with R & T crops face special constraints related to the biology, genetics and agronomy of these crops. Approaches that work well in diploid seed-propagated crops are more difficult to apply to vegetatively propagated polyploidy crops, such as potato and sweetpotato. On the other hand, the benefits of transgenic modifications may be greater for these crops precisely because of the difficulty of conventional breeding. Potato was among the first food crops to be genetically engineered about 15 years ago, with genes conferring resistance to viruses (Ooms *et al.* 1987; Stokhaus *et al.* 1987; Lawson *et al.* 1990). Since then, many other genes have been introduced and commercial varieties with resistance to potato viruses and insects have been released (Ghislain *et al.* 1999; James 2001). However, fearing negative reactions of consumers to transgenic technology, processed-food retailers have announced their decision to stop using genetically modified (GM) potatoes in their products until these have higher public acceptance. As a result, varieties released in the US are no longer available at the commercial scale.

Soon after potato, sweetpotato was genetically transformed at CIP in 1985 (Dodds *et al.* 1987). Since then, transformation technology has been applied in sweetpotato to introduce resistance to sweetpotato weevil (Zhang *et al.* 2000) and sweetpotato feathery mottle virus (SPFMV) (Okada *et al.* 2001). Cassava transformation was achieved in 1993, with CIAT being the pioneer, although the first peer-reviewed reports were published later (Li *et al.* 1996; Raemakers *et al.* 1996; Schöpke *et al.* 1996).

All three major R&T crops can be genetically engineered using *Agrobacterium tumefaciens*-mediated transformation, which presents significant advantages over other methods, particularly in terms of low technical difficulty and affordability. As transgenics provides the only means to add new traits to existing varieties of heterozygous polyploid crops without other changes, they stand to benefit greatly from transgenic improvement. Even when transfer of traits via sexual crosses is feasible, single gene transfer is likely to be more efficient than traditional breeding. In this sense, gene transfer can be considered a very valuable complement to conventional breeding. Interestingly, issues concerning gene flow from transgenics to other cultivated crops and wild relatives are more easily handled with clonally propagated species. For example, male sterility, which may be needed when transgene introgression is undesirable, does not impede the propagation of R&T crops by farmers.

8.5 Primary traits for R&T crop improvement

Given the limited resources available for international agricultural research and development, priorities for R&T crop improvement must be selected according to scientific promise, geographic coverage, potential impact on poverty, anticipated economic and environmental benefits, and the expected level of adoption in targeted and spillover countries. A global perspective on resolving constraints and realizing opportunities for R&T crop production, protection and use would prioritize two globally important breeding targets: durable disease resistance, and carbohydrate metabolism. A consideration of useful characteristics that R&T crops could conceivably contribute to other crops by way of molecular biology would include tolerance to abiotic stresses including drought and poor, acid soils (see Table 8.1).

8.5.1 Potato late blight disease; mapping and engineering progress

The development of durable resistance to late blight disease caused by *Phytophthora infestans* (Mont.) de Bary is one of the highest priorities in potato breeding programmes and probably the most challenging endeavour of molecular breeding for the crop. The need for resistant cultivars is particularly urgent because of more severe outbreaks of the disease associated with new more aggressive genotypes of the pathogen that are resistant to commonly used fungicides (Deahl *et al.* 1991; Spielman *et al.* 1991; Fry & Goodwin 1997a, b).

Two types of resistance to this disease are commonly used in potato breeding. One is a qualitative reaction, expressed as incompatibility with specific races of the pathogen. Any of a series of *R* genes in the host can result in a hypersensitive response (HR) that contains isolates of the pathogen with matching avirulence genes

Table 8.1 Major constraints and aspirations for the improvement of three important root and tuber (R&T) crops.

	Cassava	Potato	Sweetpotato
Disease resistance	Cassava mosaic disease caused by ACMV, EACMV, EACMV-UdV and SACMV (geminiviruses); bacterial blight (*Xanthomonas axonopodis* pv. *manihotis*)	Late blight (*Phytophthora infestans*); PVY (Potyvirus); PLRV (Luteovirus); PVX (Potexvirus); bacterial wilt (*Ralstonia solanacearum*)	Sweetpotato virus disease complex caused by SPCSV (Crinivirus) and SPFMV (Potyvirus)
Carbohydrate metabolism	High dry matter/starch yield; prolonged post-harvest storage period	High dry matter/starch yield; low, stable reducing sugar contents	High dry matter/ starch yield

ACMV= African cassava mosaic virus; EACMV= East African cassava mosaic virus; EACMV-UdV = East African cassava mosaic virus Uganda variant; SACMV= South African cassava mosaic virus; PVY= potato virus Y; PLRV= potato leaf roll virus; PVX= potato virus X; SPCSV= sweetpotato chlorotic stunt virus; SPFMV= sweetpotato feathery mottle virus.

and completely prevents disease development. Eleven *R* genes from the wild species *Solanum demissum* have been used since the early days of late blight resistance breeding (Malcolmson & Black 1966; Wastie 1991). Unfortunately, these have all been overcome ('defeated') by compatible isolates of the pathogen that carry virulence (*avr*) rather than avirulence (*Avr*) alleles at matching loci (Umaerus & Umaerus 1994). *P. infestans* pathotypes compatible with most *R* genes are now common, with the most complex races in Mexico having all 11 virulence factors (Rivera Peña 1990). The second type of host plant resistance to late blight is variously referred to as partial, field, quantitative or horizontal resistance. It is observed as a rate-reducing type of resistance expressed in compatible host–pathogen interactions. Partial resistance is thought to be more durable (Simmonds & Wastie 1987; Umaerus & Umaerus, 1994) than complete resistance, as it appears to be race non-specific and based on more than a single gene. The precise genetic nature of partial resistance is not yet known.

Qualitative and quantitative resistances often co-exist in potato (Tooley 1990). Because effects of *R* genes can interfere with the detection of genes for quantitative resistance, some breeders advocate eliminating *R* genes from breeding populations (Wastie 1991; Turkensteen 1993; Colon *et al.* 1995). A complementary strategy is to expose breeding populations to isolates of the pathogen with diverse or complex virulence structures that match any *R* genes present, to permit selection for underlying quantitative resistance. Whether one adopts an '*R* gene-free' philosophy or takes the approach of pyramiding *R* genes (Mohan *et al.* 1997) and other types of different resistance factors, it is helpful to know the genomic location of *R* genes, and the availability of selectable markers for them would be beneficial. Five of the 11 known *R* genes have been mapped in potato: *R1* is located on chromosome 5 (Leonards-Schippers *et al.* 1992); *R2* on chromosome 4 (Li *et al.* 1998); and *R3, R6* and *R7* on chromosome 11 (El-Kharbotly *et al.* 1994, 1996).

Leonards-Schippers *et al.* (1994) found map positions for quantitative as well as qualitative resistance within the same short chromosome segment. Detecting QTL in the same chromosome segments in which specific functional genes are located may indicate that those *R* genes participate in the quantitative resistance response to *P. infestans* as hypothesized by these authors, but this finding would also be compatible with the theory that defeated *R* genes have 'residual' effects that reduce disease (Ordoñez *et al.* 1998). In line with this hypothesis are the observations that in the compatible interaction, including partial resistance, some cells undergo HR but the hyphae can escape and continue to colonize the tissue so that the pathogen is not contained (Kamoun *et al.* 1999; Vleeshouwers *et al.* 2000). However, inheritance of *R* genes and their expression has been puzzling in several cases, and this has led to a new hypothesis of interfering factors such as suppressors (El-Kharbotly *et al.* 1996; Ordoñez *et al.* 1997). Major genetic effects that appear to confer broad-spectrum resistance have been reported in *Solanum bulbocastanum* (Naess *et al.* 2000) and *Solanum hougasii* (C. Brown, pers. commun.).

Molecular markers were associated with quantitative resistance in diploid and tetraploid populations of *S. tuberosum* cultivated potato, some of which overlapped

with previously mapped *R* genes (Leonards-Schippers *et al.* 1994; Meyer *et al.* 1998; Collins *et al.* 1999; Oberhagemann *et al.* 1999). Using the cultivated *S. phureja* source of resistance reported by Cañizares and Forbes (1995), Ghislain *et al.* (2001) and Trognitz *et al.* (2001) have studied a diploid interspecific progeny (*S. phureja* × *S. tuberosum* ssp. *tuberosum, PD*) to identify a series of DNA markers that consistently segregate with resistance in the field, greenhouse and laboratory. Sandbrink *et al.* (2000) reported progress on mapping genetic loci involved in partial resistance to late blight from *S. microdontum*. Ewing *et al.* (2000) mapped both qualitative and quantitative factors for late blight resistance in a backcross progeny of *S. berthaultii* and *S. tuberosum*. They developed a multiple QTL model that explains over 65% of variation for quantitative resistance, and mapped a putatively new *R* gene to potato chromosome 10. Evidence from field trials in Peru suggests that, following challenge by an incompatible race, this *R* gene provides a degree of protection from disease caused by compatible isolates.

So far, molecular genetic insights into late blight resistance have scarcely been used in applied breeding. A major difficulty has been the lack of a unified definition of non-*R* gene-based resistance. Indeed, a literature review reveals that several types of resistance are confounded, between general resistance mechanisms due to the expression of pathogen defence proteins, escape mechanisms such as early bulking and late foliage senescence, and physiological response such as foliage or plant vigour. The variability within field repetitions is usually low; however, the variability between different locations and years is high, indicating differential effects of pathogen populations, environmental or plant physiological conditions. Nevertheless, when genetic maps of late blight resistance are overlaid, consensus regions can be identified (Figure 8.3). Reports from other species may also contribute to understanding resistance in this important plant family. For example, efforts to map late blight resistance factors are ongoing in pepper (Pflieger *et al.* 1998) and tomato (Jones & St Clair 1998).

A review of the genomic regions of potato that have been implicated in late blight resistance identifies 13 segments that contain 81% of all QTL detected in the relatively large number of sources evaluated to date (Figure 8.3). These regions have logically been proposed as deserving special interest for both gene cloning and marker-assisted selection (Simko 2002). A breeding strategy assisted by markers flanking these consensus regions, and additional markers for 'private' QTL from potential complementary genetic resources, may well be more effective than a recurrent selection strategy that seeks to increase levels of resistance in a closed population. To this end, CIP has made crosses between diploid genotypes carrying positive QTL and susceptible tetraploid varieties. Progenies are currently undergoing field trials and marker association studies to validate and compile resistance QTL. Molecular diversity studies employing candidate genes, comparative mapping and expression assays are aimed at identifying and combining complementary resistance sources in the context of a base-broadening breeding programme.

Soon after *R* genes were mapped in potato, map-based cloning began and produced its first success with the isolation of the *R1* gene (Ballvora *et al.* 2002).

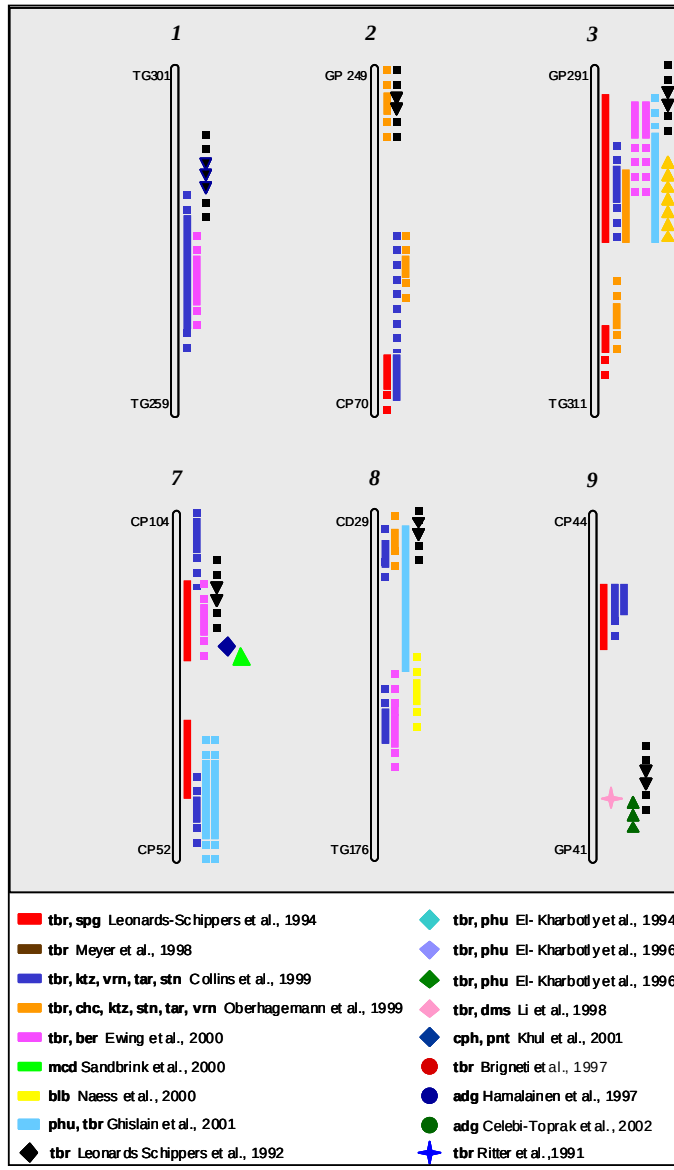
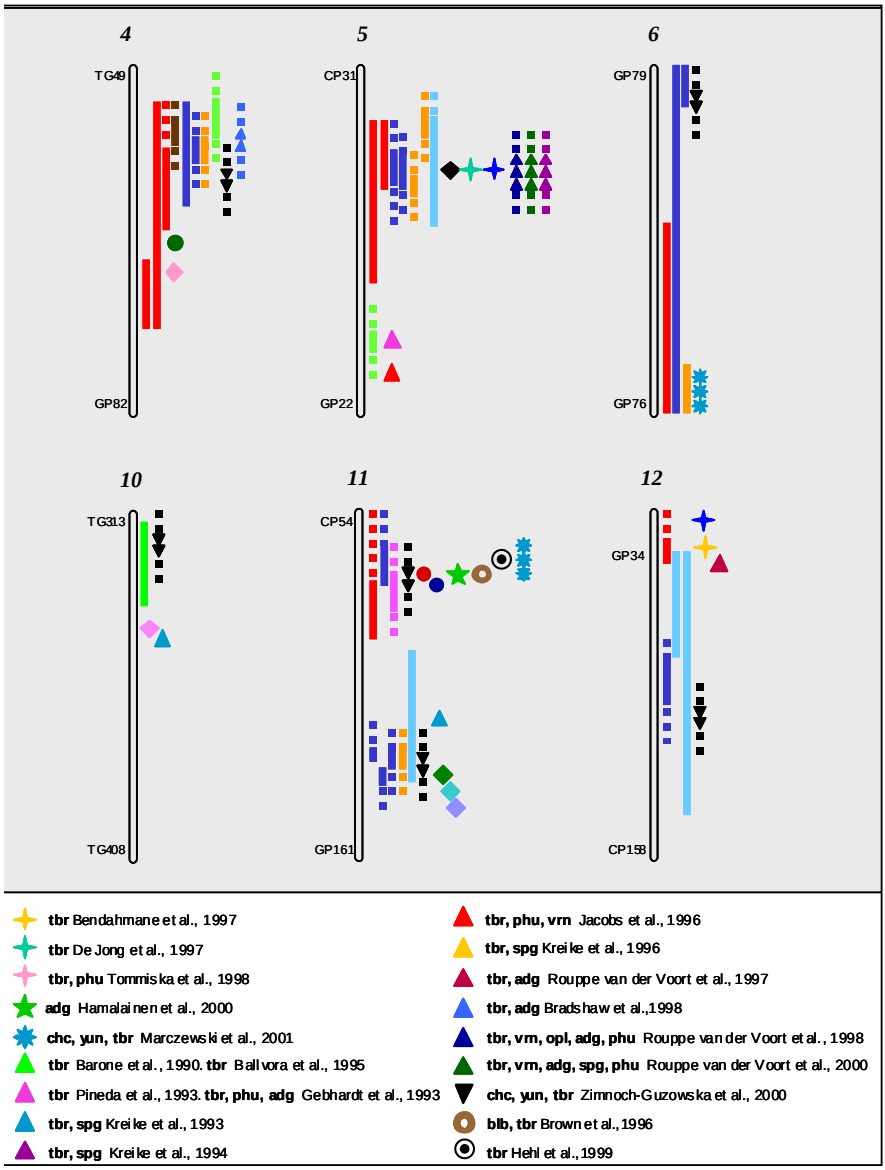


Figure 8.3 Consensus map of resistance and defence regions in potato. Positions are approximated on the basis of common or overlapping sets of RFLP and SSR anchor markers reported in 38 publications. Symbols to the right of chromosomes 1–12 classify QTL and resistance genes by target pest or pathogen: *Phytophthora* sp. QTL (━); *Phytophthora* sp. R genes (◆); PVY (●); PVX (✦); PVA (✱); PLRV (✳); *Globodera* sp. *R* genes (▲); *Globodera* sp. QTL (▲); *Erwinia* sp. QTL



(▼); *Meloidogyne* sp. (⊙); *Synchitrium* sp. (⦿). Tints of symbols refer to the respective publications, cited below map figure. Names to the left indicate reference markers (normal font), after Gebhardt *et al.* (1991) and Tanksley *et al.* (1992). QTL indicated with continuous symbols were published with flanking markers; discontinuous symbols place QTL at best anchor marker.

Positional cloning is being applied to the isolation of *R* genes from two disease resistance hotspots, and one apparent broader-spectrum resistance region in potato. Two cases so far (Ballvora *et al.* 2002; B. Baker, pers. commun.) have revealed tandemly arranged members of the nucleotide-binding, leucine-rich repeat family of resistance genes, like those previously demonstrated to confer isolate-specific resistance to viruses, bacteria, fungi and nematodes (Hammond-Kosack & Jones 1997). Improved knowledge of resistance gene organization, structure and function and new opportunities for engineering resistance and pyramiding resistance genes will no doubt follow rapidly.

Hypersensitive resistance has been engineered in other systems through either pathogen-induced or constitutive expression of *R* genes to give rise to cell death and activation of general defence mechanisms (Hammond-Kosack *et al.* 1994; Strittmatter *et al.* 1995). However, limited scope for practical use of these approaches has emerged. Hypersensitive resistance can also be activated by mutations or mediated by transgene expression (Mittler & Rizhsky 2000). The resulting plants with a phenotype referred to as lesion-mimic show enhanced resistance to pathogens due to the activation of the hypersensitive reaction. However, constitutive expression of the lesion-mimic may well impose a yield penalty. An alternative path is to develop inducible expression of this phenotype dependent on the presence of the pathogen or on a decision by the farmer to apply a chemical induction system. No such systems are available for practical use at present.

A number of new antimicrobial proteins have been shown to affect fungal disease development in potato. Recently, the expression of a synthetic cationic antimicrobial peptide gene derived from the insect proteins cecropin and melittin has also been shown to confer resistance to several phytopathogens into transgenic potato (Osusky *et al.* 2000), and the expression of an alfalfa antifungal peptide in potato has also been shown effectively to control a fungal pathogen (Gao *et al.* 2000). These results indicate progress in the discovery and engineering of disease resistance to pathogens, but there has been no further report of the effectiveness of these transgenic plants in controlling the late blight disease. Additionally, serious food safety concerns will be raised about synthetic peptides unless these are strictly expressed in non-edible tissues of the plant. An approach that has not been given enough attention is to pyramid antimicrobial protein genes in susceptible potato. Synergistic effects may be observed and result in valuable levels of disease resistance even from proteins that are not very efficient individually in conferring resistance to diseases. Research at CIP is pyramiding lysozyme, glucanase and osmotin genes, while another group in Argentina is testing the combined effects of ribosome-inactivating proteins, chitinases and other antimicrobial proteins against late blight (Mentaberry *et al.* 1997).

8.5.2 Contemporary approaches to durable resistance to potato late blight

Disease-related – or, more generally, stress-related – genes are likely to be good candidates for comparative genomics across greater taxonomic distances because there is a pressure for a coordinated and more-or-less immediate response to extraor-

dinary conditions; whereas carbohydrate metabolism-related gene positions may be less predictable as carbohydrates are usually buffered (stored) and thus there should be less need for immediate up- or down-regulation. Horizontal resistance to late blight in potato is thereby a logical trait to benefit from applying available comparative genetic and genomic information from a range of close and distant sources.

Research on late blight resistance at CIP involves a series of cultivated and wild species, diploid mapping progenies, genebank accessions, and selected progeny from the tetraploid breeding programme. Three complementary approaches are being combined to identify and compare genes and genome regions involved in late blight resistance and to link sequence information to biological function:

- Trait-based, through attempts to directly associate DNA polymorphism at candidate gene loci with resistance phenotypes.
- Gene function-based mapping and synteny assessment with conserved orthologous markers from related species.
- Gene expression-based, which provides an unbiased approach to implicating and identifying genes in the host–pathogen interaction.

8.5.2.1 Candidate gene associations with resistance in mapping populations

Late blight resistance is most likely not a unique mechanism, and must at least partially draw on the same genes that confer resistance to other pathogens. A rich literature exists on the classical and molecular genetics of a wide range of host–pathogen interactions. In particular, molecular information on genes involved in pathogen recognition (*R* genes) and general defence response genes can be used to query phenotypic variants/segregates. This knowledge led us to test the association between QTL for resistance to late blight and genes known to be involved in other pathosystems, or expected to play a role in resistance defence pathways.

In collaboration with Kansas State University and the International Rice Research Institute, CIP developed a collection of 120 primers corresponding to 29 defence genes and *R* gene motifs. Markers from this effort include consensus degenerate primers amplifying conserved regions of cloned resistance genes (*R* gene analogues) and probes and primers corresponding to genes known or suspected to play a role in plant defence.

Association of candidate genes with a number of QTL in CIP's PD population (*S. phureja* × *S. tuberosum* ssp. *tuberosum* mapping population) provided evidence in favour of their possible role in late blight resistance (Trognitz *et al.* 2002). The strongest associations were found between resistance QTL on chromosomes 3 and 12, with copies from the WRKY transcription factor family and PAL (a key enzyme in the phenylpropanoid pathway). Weaker associations were found between defence-related candidate genes belonging to the groups of signal transduction, pathogenesis and stress-related genes. The scope of such findings is mitigated by two considerations: first, the assumption that candidate genes with no polymorphism in the test population are randomly distributed among all classes of genes tested; and second, the fact that co-localization of a candidate gene and a QTL does not necessarily

imply genetic determination. The hypothesis that quantitative resistance can be largely explained by known *R* gene analogues was not supported for the analysed populations, as only 4% of *R* gene analogues detected were associated with QTL for resistance. Further, some major QTL for quantitative resistance were not found to be associated with any of the candidate genes used in our studies.

Additional studies revealing that a candidate gene has an expression pattern that is coincident with the onset and development of the disease, or differential in resistant and susceptible progeny, would provide complementary evidence in favour of the involvement of particular genes. However, even expression analyses are not conclusive in implicating or ruling out candidate genes, since allele function may be based on differential biochemical properties at the protein level. Moreover, expression patterns are prone to artefacts and errors. Hence, genetic transformation or expression interference will eventually be needed to demonstrate the precise role of candidate genes associated with QTL.

8.5.2.2 Use of functional marker loci (genes) to enhance comparative mapping

Comparative mapping is concerned with comparing locations of genes in one species with locations of the same genes in other species. Until recently, this comparison took place between closely related species; however, new advances allow comparisons across wider taxonomic groups (Fulton *et al.* 2002). Insights into the level of linkage conservation between a well-researched species such as *Arabidopsis* and a less-documented one such as potato will help to further locate genes and improve functional maps of under-researched species.

Tomato EST sequences from a series of core libraries were compared at Cornell University against the DNA sequences of the *Arabidopsis* genome (Fulton *et al.* 2002). As it pertains to a close relative, the database housing this tomato information [available via the Solanaceae Genomics Network (SGN) web-site (http://www.sgn.cornell.edu)] should be well suited for purposes of comparative mapping in potato. In more detail, about 1000 matches of the tomato EST were found with *Arabidopsis* sequences at a high level of confidence (E-value $1e^{-15}$), said to comprise a 'conserved orthologous set' (COS). About 500 of these sequences were initially mapped in an interspecific F_2 tomato population. Of the first 400 COS to be mapped, 188 were assigned functional annotations including a functional classification based on the Munich Information Center for Protein Sequences (MIPS) functional catalogue (for web-site, see References).

At CIP, 49 COS were selected from the mapped and functionally annotated subset for analysis with respect to possible relationship with horizontal resistance to late blight. Taking into account that resistance-related genes are likely to occur in clusters, the selection was also based on the functional annotation of the COS. The broad categories of COS selected were: pathogen recognition/elicitor inducible; signal transduction; stress- or metabolism-related; ion channel-related; and carbohydrate/photosynthesis-related genes. Additional selection criteria were single copy, genome representation, restriction enzyme use and co-location with previously determined resistance QTL for the trait in potato. Potato positions were predicted from tomato,

based on known syntenic relations. For even distribution, four COS were selected from each homeologous potato-tomato chromosome, while manually counterbalancing gene function, co-location with late blight QTL, even distribution, and minimizing the use of different restriction enzymes for RFLP (in tomato). RFLP was used as a conservative means for genetic transfer of the first group of COS to potato, while PCR COS are being developed for practical use. About 10 COS (~20%) showed polymorphism (using either PCR or CAPS) in at least one of the two mapping populations; five markers (~10%) were polymorphic in both populations. Only one COS mapped in both populations at the expected position (on chromosome 8). Further experimental validation is underway.

For the design of PCR primers as a more practical tool for use of the COS in potato, the tomato EST sequences were compared to sequences from the potato gene index (StGI) at The Institute for Genomic Research (TIGR) to obtain the corresponding potato EST sequences. PCR primers were designed using Primer3 (Rozen & Skaletsky 1996) for target lengths of 600 bp using otherwise standard parameter values of Primer3. Primer COS are mapped in each of two diploid reference populations in use at CIP, BCT (*S. berthaultii* and *S. tuberosum* mapping population; Bonierbale *et al.* 1992; Ewing *et al.* 2000) and PD (*S. phureja* × *S. tuberosum* ssp. *tuberosum* mapping population; Ghislain *et al.* 2001), using subsets of individuals from respective original progenies indicated as most informative by MapPop procedures (Vision *et al.* 2000). QTL mapping of late blight resistance has previously been performed in both populations. To date, five out of seven RFLP markers assayed in potato have mapped to the expected positions in the two populations. Most COS showed several polymorphic PCR products. For the PCR work under way, four markers have been mapped, three of them in approximately the same positions as in tomato (Simon *et al.* 2002). In general, it could be shown that using comparative information from a closely related species at least some EST/gene-based marker can be mapped, even though these markers are highly conserved. The usefulness of the approach remains to be confirmed in other populations, particularly to determine whether the COS-based markers have a high predictive power for involvement in shared traits across species. This approach may provide ideas on how best to take advantage of comparative genome information for the more efficient discovery and use of gene diversity. Practical experience from this example also demonstrates the need for better integration of databases and tools, to minimize the need for custom tools that are prone to error for simple data transfer/conversion.

8.5.2.3 Identification of up- and down-regulated genes as candidate determinants of QTL

Gene expression studies involving either compatible or incompatible late blight reactions expressed in stocks carrying qualitative or quantitative resistance will allow analysis of the metabolic changes that occur during recognitions, infection and disease progress, potentially revealing molecular genetic control strategies that can be applied to combat the disease. A study conducted at the SCRI has begun to identify the different classes of genes induced during the incompatible reaction (HR) with

the late blight pathogen (Birch *et al.* 1999). The experimental design allowed the identification of genes induced 24 hours after the inoculation of *P. infestans* under greenhouse conditions. Out of a group of 100 partial cDNA sequences, 13 were related to defence pathways, stress or senescence, while other plant sequences were apparently related to cell death. A similar study of the compatible reaction (disease) used suppressive subtractive hybridization (Beyer *et al.* 2001). This study identified genes coding for general pathogenesis-related proteins, and in addition to these genes for transcription, detoxification/transport, signal transduction and general metabolism as differentially expressed during disease progress.

CIP has also undertaken characterization of the compatible reaction by contrasting gene expression in progeny of its PD mapping population with different levels of quantitative resistance. The approach seeks to identify genes that are regulated in response to disease development, and that map to a strong QTL on chromosome 12 (Ghislain *et al.* 2001; Trognitz *et al.* 2002). Pools of expressed sequences from the four possible genotypic classes were assayed for differences during disease development under field conditions in Peru, and under greenhouse conditions at the CRP-GL in Luxembourg. Differentially expressed sequences were detected by cDNA-amplified fragment length polymorphism (AFLP). Preliminary conclusions of this study coincide with the previous ones, in that stress- and senescence-related genes were differentially expressed among the genotypic classes and time series samples (D. Evers, pers. commun.). The studies also revealed apparent differential expression of genes involved in carbohydrate metabolism. The confirmation of map position of such genes is an essential next step to investigation of the role of these new candidate genes as determinants of the QTL in question.

The availability of potato microarrays is opening the way to larger, more integral analysis of gene expression. A unique set of proteinase inhibitors has been found to be induced early in the late blight infection process, but are later repressed (C. Smart, pers. commun.). Sub-arrays comprising putative defence response and signalling genes have also been prepared from tomato ESTs, and recently applied to identifying genes expressed under different types of compatible reactions between *Phytophthora infestans* and tomato (Smart *et al.* 2002). Repetitive probing of microarrays with RNA from contrasting or differentially challenged plants and from critical time periods will unravel complex gene expression patterns that will need new bioinformatic tools to be interpreted (for a review, see Reymond 2001). Such findings are already providing new information about the physiology of late blight disease and will likely lead to better understanding of mechanisms of tolerance. The outcome of such approaches will be to tag genes whose patterns of expression match disease development, resistance reactions and tissue-specific differential responses for manipulation in breeding programmes. The expectation that genes underlying horizontal resistance interact through a balanced gene network, implying complex expression patterns as opposed to simple on/off patterns, makes a strong case for applying more system-oriented analytical methods like metabolic control analysis

(MCA; pioneered by Kacser & Burns 1973) or biochemical systems analysis (BSA; Savageau 1969) in the future.

8.6 Molecular approaches to characterizing and improving carbohydrate metabolism in R&T crops

Starch production for food and non-food use is one of the most important agro-industries worldwide with a total world production of 37 millions tons of which nearly 30% is from R&T crops (Jesch 1996). Starch demand is influenced by its versatility. Functional properties of starches are of direct technical and economic interest to competitive food and non-food industries. Starch is a very versatile raw material because it can be modified (either physically, chemically or enzymatically) from its native form by relatively simple processes. The industrial modification of native starches has expanded the range of potential uses by tailoring particular products to specific industrial needs. The use of artificially modified starches has raised a negative connotation particularly within the food industry, as a consequence of potential collateral effects (i.e. cancerous products). The possibility of genetically modifying starches could not only solve this problem, but also reduce the cost of the final product since it will avoid or greatly reduce the need for artificial manipulation.

Starches from potato, sweetpotato and cassava have large genotypic variation in terms of their functional properties (Collado & Corke 1997, 1999; Khalil & Henry 1997). Particular physical and chemical properties of individual starches are the keys to their commercial success. The yield of R&T crops is considered to be largely a function of sink strength and sink activity of the respective storage organs (Kays 1985). Productivity and quality represent the outcome of complex interrelated carbohydrate metabolism traits involving cross-talks among networks of internal and external stimuli, biochemical pathways and genetic components. A range of physiological processes including resource accumulation, partitioning, tuberization or root thickening, bulking, storability and mobilization of reserves all influence R&T crop productivity, quality and value.

Sink strength of R&T crops in large part depends on the mechanism for the transfer of nutrients away from the transport system and the physical and biochemical isolation of the transported carbon in the storage organ. Metabolism provides the energy for the unloading of assimilates transported to the storage root via the phloem and the synthesis of starch and all the carbon skeletons necessary for cell growth. Therefore, phenotypic variations of starch productivity and functional properties of R&T crops are often attributed to the intra-specific difference in sugar and starch metabolism, which are genetically determined but also subject to influence by environmental factors such as soil temperature, humidity, light and soil fertility.

Molecular tools are essentially complementary to conventional breeding for the improvement of starch yield and quality of R&T crops. Potato accounts for the great majority of the research carried out to date. However, the knowledge derived from

potato will benefit cassava and sweetpotato since the three crops share similar physiological mechanisms and biochemical pathways in sugar and starch metabolism.

8.6.1 Molecular manipulation of starch functional properties through expression of transgenes

A good example of creating novel variation of starch quality is the change of amylose/amylopectin ratio in R&T crops. Starch is a mixture of two structurally different polymers, amylose and amylopectin. Amylose is linear and amylopectin is highly branched – and each plays a critical role in the ultimate functionality of native starch and its derivatives. Viscosity, shear resistance, gelatinization, texture, solubility, tackiness, gel stability, cold swelling and retrogradation all depend at least in part on amylose/amylopectin ratio.

Using antisense technology, amylose-free potato and sweetpotato were successfully created through inhibiting granule-bound starch synthase gene expression (GBSS) (Visser *et al.* 1991; Kuipers *et al.* 1991, 1994a, b; Wang ShuJen *et al.* 1999; Wang *et al.* 2001; Kimura *et al.* 2001). Transgenic potato plants with complete inhibition produced amylose-free starch (Visser *et al.* 1991). Expression of the antisense GBSS gene using the GBSS promoter resulted in a higher stability for amylose-free trait than with the CaMV 35S promoter. Field analyses of the transgenic clones indicated that inhibition of GBSS gene expression could be achieved without significantly affecting the starch and sugar content of transgenic tubers (Kuipers *et al.* 1991, 1994b). Kimura *et al.* (2001) identified amylose-free sweetpotato after screening 26 transgenic plants carrying an antisense GBSS sequence.

Besides the change of amylose/amylopectin ratio, modification of branching of amylopectin by molecular manipulation has also been explored. Kortstee and Vermeesch (1996) reported that expression of *E. coli* branching enzyme *glgB* gene in tubers of amylose-free transgenic potatoes leads to an increased branching degree of amylopectin. A 25% increase in branching points of amylopectin was reported. Lloyd *et al.* (1999) reported that simultaneous antisense inhibition of two starch-synthase isoforms in potato tubers leads to accumulation of grossly modified amylopectin. Safford *et al.* (1998) generated transgenic plants with inhibition of starch branching enzyme by expressing antisense constructs containing cDNAs for potato starch branching enzyme (SBE). A consistent alteration in starch gelatinization properties was only observed when the level of SBE activity was reduced to below approximately 5% of that of the wild genotype.

8.6.2 Starch grain formation and morphology

The ability to alter the size and shape of starch granules has significant implications for specific industrial utilization of R&T starches. For instance, the cosmetic industry requires starches with uniformly small granule size. Starch granule shape and surface are critical factors when starch is used as a surface carrier of colours, flavours and seasonings. Tauberger *et al.* (1999) showed that the size of starch grains was re-

duced in transgenic potato expressing a transgene of invertase and an antisense gene of ADP-glucose pyrophosphorylase (AGPase). Geigenberger *et al.* (2001) developed transgenic potato plants with altered plastidic adenylated (ATP/ADP) transporter activity. The potato tubers with antisense inhibition of the ATP/ADP transporter contained 50% reduced levels of ADP-glucose, whereas transgenic tubers over-expressing the gene contained up to two-fold increased levels of ADP-glucose. Starch grains in antisense tubers were 50% smaller than those of the wild type. Over-expression of the ATP/ADP transporter resulted in larger starch grains with more angular shape. The results suggested that ADP-glucose pyrophosphorylase function *in vivo* is limited by the supply of ATP.

8.6.3 Starch-sugar conversion

The rate and degree of conversion of starch to sugars determines the acceptability of R&T crops as starch sources and as raw material for use in processing for the fast food industry. Potato plants transformed with a thermostable alpha-amylase (*Bacillus stearothermophilus*) and glucose isomerase (*Thermus thermophilus*) genes produce potatoes that can become sweet when warmed to 65ºC for 45 minutes (Beaujean *et al.* 2000). Starch in these transgenic potatoes was degraded to glucose by the thermostable amylase, and then further partially converted to fructose – a much sweeter sugar than glucose – by the thermostable glucose isomerase. This approach of producing transgenic enzymes in plants and activating them when needed at high temperature opened a new strategy for the starch processing industry.

Transgenic manipulation of carbohydrate metabolism genes can also dramatically reduce the severity degree of cold sweetening of potatoes after harvest. Most potatoes accumulate hexose (glucose and fructose) during storage under low temperatures, in a phenomenon called low-temperature or cold-sweetening. A reducing sugar content of 0.1% fresh weight or less is ideal for the manufacture of fried potato chips, and higher than 0.33% is unacceptable (Dale & Mackay 1994). In order to produce high-quality potato chips and French fries, growers and handlers for the processing industry have to employ expensive storage methods because most of the current cultivars are sensitive to cold-sweetening. Potatoes for processing have to be stored under controlled warm conditions during winter, which is expensive, and could induce sprouting and aging.

Transgenic plants over-expressing a mutated *E. coli* AGPase gene *glgC16* increased starch synthesis rate, and the transgenic tubers showed reduced accumulation of hexose, possibly due to an increased rate of starch synthesis (Stark & McCann 1992). However, the same increase in starch synthesis was not detected when over-expressing the same gene under a patatin promoter in potato in spite of a four- to five-fold increase of the enzymatic activity of AGPase (Sweetlove *et al.* 1996). Successful reduction of hexose accumulation was reported by antisense inhibition of *R1* genes in potato (Lorberth *et al.* 1998). *R1* encodes a protein capable of introducing phosphate into starch-like glucans. Antisense *R1* potatoes produced reduced-phosphate starch and exhibited lower starch degradation. They were consequently more resistant to

cold-sweetening compared to wild-type potatoes. Commercialization of anti-*R1* technology for potato quality may be limited because of its concomitant alteration of starch quality (Greiner *et al.* 1999).

Given that acid invertase activity is largely responsible for the low-temperature accumulation of reducing sugars (Davies *et al.* 1989), another approach to inhibiting cold-sweetening is to inhibit this key enzyme of sucrose metabolism pathways. Transgenic potatoes with an acid invertase gene in antisense direction (Zrenner *et al.* 1996), or with an invertase inhibitor gene (Greiner *et al.* 1999), have shown decreased reducing sugar contents. A reduction of greater than 90% in invertase activity was required for significant effect on reducing sugar accumulation. Through using ectopic expression of a tobacco invertase inhibitor, Greiner *et al.* (1999) decreased cold-induced hexose accumulation up to 75% without affecting tuber yield. However, the levels of hexose in the invertase inhibitor transgenic plants were still in excess of what is commercially acceptable for production of potato chips and French fries (Sowokinos 2001a).

Sucrose phosphate synthase (SPS) is responsible for combining UDP-glucose with fructose 6-phosphate to produce sucrose 6-phosphate (Suc-6-P), the precursor of sucrose, in higher plants (Huber & Huber 1992). Therefore, SPS is a key enzyme for sucrose synthesis in both photosynthetic and non-photosynthetic organs (Huber & Huber 1996). Sucrose-phosphate phosphatase immediately converts Suc-6-P to free sucrose. A 70 to 80% decrease of SPS activity in potatoes transformed with antisense SPS gene resulted in a decrease of soluble sugars (Krause *et al.* 1998).

8.6.4 Starch content

Efficient accumulation of starch is a primary determinant of both yield and quality in R&T crops. Inhibition of key enzymes involved in starch synthesis such as sucrose synthase (Zrenner *et al.* 1995) and AGPase (Muller-Rober *et al.* 1992) have been shown to abolished potato starch synthesis, while over-expression of AGPase increased starch synthesis rate (Stark & McCann 1992). A 1.5- to 3-fold reduction of starch content was reported in cassava transgenic plants expressing a small subunit gene of AGPase in an antisense orientation (Munyikwa *et al.* 2001). Inhibition of cytosolic phosphorylase in potato did not noticeably alter dry matter content (Duwenig *et al.* 1997). Sweetpotato AGPase genes have been cloned (Bae & Liu 1997), but there is no report on transgenic manipulation of AGPase in sweetpotato. Tobias *et al.* (1999) reported that over-expression of SPS activity in source leaves of potato plants under the control of a modified (CaMV) 35S promoter did not significantly alter AGPase activity, but increased tuber dry matter content to 20%. Molecular genetic manipulation can abolish starch synthesis by inhibiting key enzymes involved in starch synthesis; however, the increase of starch content by over-expressing key enzymes was much less apparent. This might be due to the limitation of photoassimilate sucrose supply to the sink, or to an increase of starch turnover by an unknown mechanism. Munyikwa *et al.* (1997) have summarized the results of cloning and characterization of the genes coding for the main enzymes involved in cassava starch

biosynthesis. These are the genes coding for the ADP glucose pyrophosphorylase B and S subunits, branching enzyme, GBSS and their isoforms.

8.6.5 Productivity and sink strength

Carbohydrate productivity is a net accumulation of photoassimilates. Plant productivity is dependent on the effectiveness of photosynthesis, sucrose transport and starch accumulation. In the case of potato, sweetpotato and cassava, the efficiency of storing translocated photoassimilates depends highly on sink strength. A number of transgenic studies have demonstrated that pleiotropic effects of the modification of carbohydrate metabolism genes often affect tuber size and number. AGPase is a key enzyme for starch synthesis, and AGPase genes' expression activity is also considered as important for sink strength. Antisense inhibition of AGPase gene caused an accumulation of soluble sugars in tubers and a significant increase of the total tuber fresh weight, while decreasing dry weight of tubers (Muller-Rober *et al.* 1992). AGPase activity can be enhanced by sucrose supply. For example, exogenous injection of sucrose solution to sweetpotato roots enhanced the activity of AGPase (Tsubone *et al.* 2000; Kadowaki *et al.* 2001). These results suggested that AGPase activity should also be enhanced by imported sucrose supply from leaves.

Over-expressed cytosolic or apoplastic invertase in the wild-type or AGPase antisense background also significantly modified the size and number of tubers in various transgenic lines (Tauberger *et al.* 1999). Expression of a bacterial glucokinase, either in the wild-type or in the apoplastic invertase-expressing background, led to changes in the levels of glucose and glucose 6-phosphate in potatoes (Fernie *et al.* 2000). However, these changes had little effect on the partitioning of carbohydrate or tuber size. Increase of starch content and tuber size has also been reported through inhibition of the expression of the potato gene encoding 14-3-3 protein (Wilczynski *et al.* 1998).

Sucrose synthase is another key enzyme affecting potato sink strength. Strong activity was observed for sucrose synthase genes in potato tubers (Salanoubat & Belliard 1989; Fu & Park 1995; Hajirezaei *et al.* 2000). Antisense inhibition of sucrose synthase led to a strong accumulation of reducing sugars and an inhibition of starch accumulation in potato tubers (Zrenner *et al.* 1995). Evidence from two other root crops, namely carrot (Sturm *et al.* 1999) and radish (Usuda *et al.* 1999), also supports the correlation between sucrose synthase gene expression and sink strength.

It is notable that R&T crops can share transgene resources for the manipulation of starch metabolism. Salehuzzaman *et al.* (1993) isolated a cDNA encoding GBSS in cassava. Expressing the antisense sequence of the cassava cDNA in potato inhibited endogenous potato GBSS gene expression to various degrees. Potato plants with complete inhibition of potato GBSS gene expression by this cassava antisense GBSS gene produced starch that was nearly amylose-free. These results suggest the feasibility of using other root crop genes to manipulate carbohydrate metabolism traits of potato tubers. However, while antisense cassava GBSS gene can abolish amylose synthesis in potato tubers, over-expression of cassava GBSS (sense) could

not completely restore the amylose production in an amylose-free mutant of potato (Salehuzzaman *et al*., 1999) indicating that genes or expression cassettes from different crops have their own respective important properties.

8.6.6 Genomics contribution to the improvement of starch yields and quality in R&T crops

The above-mentioned examples of transgenic research showed the promising prospect in molecular manipulation of carbohydrate metabolism in R&T crops. Meanwhile, these examples demonstrate the complexity of the interrelated pathways, the necessity of improving the molecular characterization of all of the related subsets of the genes involved in starch metabolism, and the importance of understanding these genes' specific functions, phenotypic consequences and interactions with the environment.

Tools of genomics will greatly assist improvement of starch yield and quality by allowing precise hypothesis testing toward the development of referential frameworks and by developing methods for rapid and high-resolution identification of beneficial alleles and desired gene combinations. This assistance will open great possibility and flexibility for breeders to use molecular genetic and genomics technologies in cultivar development. For example, potato cultivars with varying levels of tolerance to cold-sweetening have been characterized by allelic genotyping of UGPase (Sowokinos *et al.* 1997, 2000; Sowokinos 2001a, b). Different isoforms of UGPase were associated with different regulatory properties in the direction of UDP-glucose formation. Most of the cold-resistant cultivars were unique in that they contained up to two UGPase isoforms slightly more basic in charge than the most common isoform in cold-sweetening-sensitive cultivars (Sowokinos 2001a).

A number of QTL for dry matter content have been reported in potato, and several carbohydrate metabolism genes have been mapped to them, indicating their possible involvement in the dry matter accumulation (Chen *et al.* 2001). Some of the candidate genes co-located with dry matter QTL are two glucan phosphorylase genes and a Rubisco gene on chromosome 2, a starch synthase gene on chromosome 3, a starch phosphorylase and a debranching enzyme gene on chromosome 9, a Rubisco activase gene (*RCA*) and apoplast invertase gene on chromosome 10, a sucrose transporter gene on chromosome 11, and a sucrose synthase gene on chromosome 12. If polymorphic in breeding populations, these genes should be useful for marker-assisted selection to increase dry matter content in potato.

In sweetpotato, acid invertase activity was reported in freshly harvested storage roots (Takahata *et al.* 1996). Sucrose synthase activity was also found in storage roots, with activity increasing during development (Saitou *et al*. 1997). Both acid invertase and sucrose synthase were found to be active in sweetpotato suspension cells (Wang *et al*. 2000). Although sweetpotato is a starch crop, and starch accumulation increases rapidly during storage root development, there is no report yet on which sucrose cleavage pathway, invertase or sucrose synthase, is more important during storage root development.

As a first step in characterizing sucrose metabolism during sweetpotato storage root enlargement, CIP and the Canadian (AAFC) Potato Research Center are investigating the activity of sucrose metabolism genes at a rapid enlargement stage, and the major enzymatic pathways for cleaving loaded sucrose to phosphoglucoses, as associated with root sink strength. Two cDNA libraries have been constructed using mRNA extracted, respectively, from fibrous roots and rapidly enlarging storage roots of a high dry matter sweetpotato variety. Random cDNA clones have been sequenced to generate ESTs. From a pilot set of 1000 ESTs, six from the storage root and one from the fibrous root were found to be homologous with sucrose synthase, and one EST from the storage root was homologous with UDP-glucose pyrophosphorylase. No ESTs of invertase or glucose phosphorylase were found in either type of roots in this first set of ESTs. The much higher expression level of the sucrose synthase genes than that of acid invertase genes suggests that the sucrose synthase-UGPase pathway plays a more crucial role in sucrose metabolism during storage root development (X.Q. Li and D.P. Zhang, unpublished results). These experiments also showed that sucrose synthase expression is associated with the sink activity and capacity in sweetpotato.

Based on the ESTs generated and other known sequences, functional molecular markers for starch metabolism in sweetpotato have been developed using cleaved amplified polymorphic sequences (CAPS). To date, 22 loci have been mapped (under a single dose condition) on a previously constructed framework linkage map of sweetpotato built with AFLP markers (Hurtado *et al.* 2001).

8.7 Future prospects

Genomics research has led to a rapid increase of available genes involved in carbohydrate metabolism from potato, sweetpotato and cassava, and to sequence and organizational information regarding host plant resistance and defence strategies. Some 60 carbohydrate metabolism genes and 19 resistance genes have been located on potato genetic maps. Genetic manipulation, either up-regulating or down-regulating expression of known genes, has led to the modification of carbohydrate metabolism in R&T crops. Appropriate modifications have improved plant productivity, tuber number, tuber size, dry matter content and starch physical-chemical properties. Novel traits can also be genetically engineered in potato plants, such as in the case of converting starch to fructose under high-temperature conditions and the production of fructan, a high molecular-weight carbohydrate polymer (Van der Meer *et al.* 1994; Rober *et al.* 1996). Genetically improving starch quality in one crop can also be achieved by using genes of another tuber or root crop. Gene regulation and transfer, comparative mapping, association genetics and genomics applied to the characterization and use of germplasm based on allele mining have significant potential for more directed improvement of the starch and resistance properties of R&T crops. Unfortunately at present, concerns and suspicions over profits to be gained from genes and germplasm are impeding their optimal characterization and

use. Both fair and practical approaches are urgently needed to permit the effective collaboration between technology-rich and diversity-rich parties.

8.7.1 Carbohydrate trait improvement

More research is required to investigate the expression and regulation of carbohydrate metabolism genes for understanding carbohydrate metabolism networks. The optimization of transgenic approaches to reduce pleiotropic effects of starch-related modifications is important for practical utility of transgenic plants in commercial applications. The resistance of potato to cold-sweetening requires further research because the reducing sugar contents of cold-stored potatoes are still too high for processing potato chips. There is also little research at the molecular level on the stability of dry matter contents under different climatic conditions. An outstanding set of Andean potato varieties is now under characterization for stability of its high dry matter and low reducing sugar contents, and it is expected that candidate gene analysis will reveal superior alleles and gene combinations that underlie the apparent advantages in this germplasm over high-yielding, improved varieties. Genetic engineering has created amylose-free potato and sweetpotato starch, but genetically creating amylopectin-free starch is still a challenge. Molecular and genomic technologies have not yet been put to practical use for improving cassava utilization. New industrial uses of cassava would be more feasible if genetic engineering of cell wall synthesis-related metabolism could facilitate the breakdown of tough-textured cassava roots. Complementary research is required to convert gene information into practical molecular markers that can be employed to characterize germplasm and assist the breeding selection of root and tuber crops. In this way, searches for variation in genes in key metabolic pathways can be used to identify loci and alleles with the best prospects for maximizing productivity and utilization traits, while minimizing negative consequences of breeding with unadapted germplasm.

8.7.2 Resistance breeding

The discovery of genes for durable resistance, other than those that are *R* gene-based, is still a major challenge. Unfortunately, the cases for the discovery of genetic determinants of QTL are still rare. In late blight resistance, most QTLs reported appear to be only of moderate genetic consequence (strength), and several chromosome regions appear to be inter-related. The use of frontier science will likely reveal increasingly precise series of gene candidates, as eventually transcriptomes will be discovered for each QTL. The use of the DNA chips and bioinformatics will allow us to narrow down from the myriad of genes within and among QTL intervals, to a smaller group of genes that would be candidates of the defensome. The recent cloning of an *R* gene for potato late blight resistance will facilitate the screening of large numbers of germplasm accessions for additional alleles at the *R1* locus, some of which may confer non-specific, or even non-host, types of resistance. At the same time, careful analysis of strategically selected germplasm accessions and mapping

populations for a series of resistance components will provide insight into whether QTL may influence different resistance factors that act on multiple targets in the pathogen. In this way, combining phenotypic and molecular characterization, the relationships between complete and partial resistance, the role of putative suppressors of *R* genes, and potentially, factors governing interactions with phenology and physiology such as plant maturity can also be unravelled, for better understanding of and ability to breed for durable resistance.

In new approaches to germplasm characterization, polymorphisms can now be assessed by the use of specific primers for the *R1* gene, and haplotypes determined. PCR products can further be digested and sequenced, and homology to predicted amino acid sequences can be compared across haplotypes and against related proteins from other sources to permit the identification of conserved and variable regions, and the development of phylogenetic inferences. Thus, allelic differences among resistance gene homologues can be catalogued, combined with phenotypic tests, and used to explain variation in gene function. In the future it is likely that resistance genes may be engineered with broader recognition spectra resulting in greater durability, and to induce response pathways that enhance defence.

8.7.3 High-throughput genotyping and association mapping

R & T crops could take advantage of the investment in technology development that has already been made by the broader community, especially by the Human Genome Project. Because of the polyploid and heterozygous nature of R&T crops, several alleles are expected to be present among the redundant sequence tags of each given gene, making the EST database a potential source for SNP discovery. While the potential quantitative accuracy of SNP genotyping is especially attractive to polyploid R&T crops, improvements to the detection technology that are robust to ploidy are still needed for practical application to sweetpotato, cassava and potato. Encouragingly, allelic discrimination through SNP detection has recently been demonstrated in tetraploid potato permitting the distinction not only between homo- and heterozygosity, but also among different heterozygous states (Rickert *et al.* 2002).

Methods for the rapid and high-resolution identification of beneficial alleles and polymorphisms, short of the development of special stocks for QTL and map-based cloning approaches, are embodied in the concept of association genetics. In association genetics, polymorphisms that correlate with phenotypic variation are identified by the evaluation of candidate gene diversity across natural populations. The approach has been used most effectively so far in human genetics, but holds great promise for gene discovery directly in genebank accessions or bred germplasm, particularly in cases for which common genetic stocks are not ideal for comparative studies. To apply these approaches based on linkage disequilibrium, greater attention will need to be given to determining the population structure of genebank materials and bred germplasm that are variable for targeted traits. Association tests in germplasm, combined with high-throughput genotyping procedures and global expression analysis such as microarray technology, appear to offer new opportuni-

ties to understand gene function and efficiently discover useful alleles for root and tuber crop breeding. Methods accounting for population structure and that reduce spurious associations between candidate gene markers and phenotypic classes (for a review, see Buckler & Thornsberry 2002) will need to be developed for the complex genetics and clonal nature of R&T germplasm collections if this strategy is to be applied to discovering useful genes in genebanks or farmers' fields.

Acknowledgement

The authors gratefully acknowledge Maria del Rosario Herrera for drawing the consensus map of resistance and defence regions in potato, Joe Tohme for his input into the orientation of research toward molecular breeding for cassava quality, Theresa Fulton and Steve Tanksley for generously providing early access to the tomato COS data, and Leticia Portal for assitance with testing the COS in potato.

References

Ablett, E., Seaton, G., Scott, K., *et al.* (2000) Analysis of grapes ESTs: global gene expression patterns in leaf and berry. *Plant Science*, **159**, 87–95.

Ahn, S. & Tanksley, S.D. (1993) Homologous relationships of rice, wheat and maize chromosomes. *Proceedings of the National Academy of Sciences, USA*, **90**, 7980–7984.

Akano, A.O., Dixon, A.G.O., Mba, C., Barrera, E. & Fregene, M. (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics*, DOI 10.1007/s00122-002-0891-7. Published online: 8 May 2002.

Alexandratos, N. (1995) The outlook for world food and agriculture to year 2010. Population and food in the early twenty-first century: meeting future food demands of an increasing population. International Food Policy Research Institute (IFPRI), Washington DC, pp. 25–48.

Asiedu, R., Wanyera, N.M., Ng, S.Y.C. & Ng, N.Q. (1997) Yams. In: *Biodiversity in Trust – Conservation and Use of Plant Genetic Resources in CGIAR Centers* (eds D. Fuccillo *et al.*), pp. 57–66. Cambridge University Press.

Bae, J.M. & Liu, J.R. (1997) Molecular cloning and characterization of two novel isoforms of the small subunit of ADPglucose pyrophosphorylase from sweetpotato. *Molecular and General Genetics*, **254**, 179–185.

Ballvora, A., Hesselbach, J., Niewohner, J., Leister, D., Salamini, F. & Gebhardt, C. (1995) Marker enrichment and high-resolution map of the segment of potato chromosome VII harbouring the nematode resistance gene *Gro1*. *Molecular and General Genetics*, **249**, 82–90 (Re Figure 8.3).

Ballvora, A., Ercolano, M.R., Weiß, J., *et al.* (2002) The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant Journal*, **30**, 361–371.

Barone, A., Ritter, E., Schachtschabel, U., Debener, T., Salamini, F. & Gebhardt, C. (1990) Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Molecular and General Genetics*, **224**, 177–182 (Re Figure 8.3).

Beaujean, A., Ducrocq-Assaf, C., Sangwan, R.S., Lilius, G., Bulow, L. & Sangwan-Norreel, B.S. (2000) Engineering direct fructose production in processed potato tubers by expressing a

bifunctional alpha-amylase/glucose isomerase gene complex. *Biotechnology and Bioengineering*, **70**, 9–16.

Bendahmane, A., Kanyuka, K. & Baulcombe, D.C. (1997) High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to potato virus X in tetraploid potato. *Theoretical and Applied Genetics*, **95**, 153–162 (Re Figure 8.3).

Bernacchi, D. & Tanksley, S.D. (1997) An interspecific backcross of *Lycopersicon esculentum x L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. *Genetics*, **147**, 861–877.

Beyer, K., Binder, A., Boller, T. & Collinge, M. (2001) Identification of potato genes induced during colonization by *Phytophthora infestans*. *Molecular Plant Pathology*, **2**, 125–134.

Birch, P.R.J., Avrova, A.O., Duncan, J.M., Lyon, G.D. & Toth, R.L. (1999) Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. *Molecular Plant-Microbe Interactions*, **12**, 356–361.

Bonierbale, M., Plaisted, R. & Tanksley, S.D. (1992) Genetic mapping and utilization of quantitative trichome-mediated insect resistance in potato. *Netherlands Journal of Plant Pathology*, **98**, Supplement 2, 211–214.

Bonierbale, M., Plaisted, R.L. & Tanksley, S.D. (1988) RFLP maps of potato and tomato based on a common set of clones reveal modes of chromosomal evolution. *Genetics*, **120**, 1095–1103.

Bradshaw, J.E., Hackett, C.A., Meyer, R.C., *et al*. (1998) Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. *Theoretical and Applied Genetics*, **97**, 202–210 (Re Figure 8.3).

Brigneti, G., Garcia-Mas, J. & Baulcombe, D.C. (1997) Molecular mapping of the potato virus Y resistance gene Ry_{sto} in potato. *Theoretical and Applied Genetics*, **94**, 198–203 (Re Figure 8.3).

Brown, C.R., Yang, C.P., Mojtahedi, H., Santo, G.S. & Masuelli, R. (1996) RFLP analysis of resistance to Columbia root-knot nematode derived from *Solanum bulbocastanum* in a BC<sub(2)> population. *Theoretical and Applied Genetics*, **92**, 572–576 (Re Figure 8.3).

Buckler, E.S. & Thornsberry, J.M. (2002) Plant molecular diversity and application to genomics. *Current Opinion in Plant Biology*, **5**, 107–111.

Cañizares, C.A. & Forbes, G.A. (1995) Foliage resistance to *Phytophthora infestans* (Mont.) de Bary in the Ecuadorian national collection of *Solanum phureja* spp. *phureja* Juz. & Buk. *Potato Research*, **38**, 3–10.

Cato, S.A., Gardner, R.C., Kent, J. & Richardson, T.E. (2001) A rapid PCR-based method for genetically mapping ESTs. *Theoretical and Applied Genetics*, **102**, 296–306.

Celebi-Toprak, F., Slack, S.A. & Jahn, M.M. (2002) A new gene, Ny_{tbr}, for hypersensitivity to potato virus Y from *Solanum tuberosum* maps to Chromosome IV. *Theoretical and Applied Genetics*, **104**, 669–674 (Re Figure 8.3).

Chen, X., Salamini, F. & Gebhardt, C. (2001) A potato molecular-function map for carbohydrate metabolism and transport. *Theoretical and Applied Genetics*, **102**, 284–295.

Collado, L.S. & Corke, H. (1997) Properties of starch noodles as affected by sweetpotato genotype. *Cereal Chemist*, **74**, 182–187.

Collado, L.S. & Corke, H. (1999) Heat-moisture treatment of sweetpotato starch. *Food Chemistry*, **65**, 339–346.

Collins, A., Milbourne, D., Ramsay, L., *et al*. (1999) QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. *Molecular Breeding*, **5**, 387–398.

Colon, L.T., Turkensteen, L.J., Prummel, W., Budding, D.J. & Hoogendoorn, J. (1995) Durable resistance to late blight (*Phytophthora infestans*) in old potato cultivars. *European Journal of Plant Pathology*, **101**, 387–397.

Conner, P.J., Brown, S.K. & Weeden, N.F. (1998) Molecular-marker analysis of quantitative traits for growth and development in juvenile apple trees. *Theoretical and Applied Genetics*, **96**, 1027–1035.

Dale, M.F.B. & Mackay, G.R. (1994) *Inheritance of Table and Processing Quality*, pp. 285–315. Centre for Agriculture and Biosciences International (CABI), Wallingford, UK.

Davies, H.V., Jefferies. R.A. & Scobie, L. (1989) Hexose accumulation in cold-stored tubers of potato (*Solanum tuberosum* L.): the effects of water stress. *Journal of Plant Physiology*, **134**, 471–475.

De Jong, W., Forsyth, A., Leister, D., Gebhardt, C. & Baulcombe, D.C. (1997) A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on Chromosome 5. *Theoretical and Applied Genetics*, **95**, 246–252 (Re Figure 8.3).

Deahl, K.L., Goth, R.W., Young, R., Sinden, S.L. & Gallegly, M.E. (1991) Occurrence of the A^2 mating type of *Phytophthora infestans* in potato fields in the United States and Canada. *American Potato Journal*, **68**, 717–725.

Dodds, J.H., Lizarraga, R., Siguenas, C., Gutierrez, R.A. & Malamud, O. (1987) Tissue culture techniques applied to potato and sweet potato improvement. CIP. 15 pp.

Duwenig, E., Steup, M., Willmitzer, L. & Kossmann, J. (1997) Antisense inhibition of cytosolic phosphorylase in potato plants (*Solanum tuberosum* L.) affects tuber sprouting and flower formation with only little impact on carbohydrate metabolism. *Plant Journal*, **12**, 323–333.

El-Kharbotly, A., Leonards-Schippers, C., Huigen, D.J., *et al.* (1994) Segregation analysis and RFLP mapping of the *R1* and *R3* alleles conferring race-specific resistance to *Phytophthora infestans* in progeny of dihaploid potato parents. *Molecular and General Genetics*, **242**, 749–754.

El-Kharbotly, A., Palomino-Sánchez, C., Salamini, F., Jacobsen, E. & Gebhardt, C. (1996) *R6* and *R7* alleles of potato conferring race-specific resistance to *Phytophthora infestans* (Mont.) de Bary identified genetic loci clustering with the *R3* locus on chromosome XI. *Theoretical and Applied Genetics*, **92**, 880–884.

El-Sharkawy, M.A. (1993) Drought tolerant cassava for Africa, Asia and Latin America. *Bioscience*, **43**, 441–451.

Ewing, E.E., Simko, I., Smart, C.D., *et al.* (2000) Genetic mapping from field tests of qualitative and quantitative resistance to *Phytophthora infestans* in a population derived from *Solanum tuberosum* and *Solanum berthaultii*. *Molecular Breeding*, **6**, 25–36.

Fernie, A.R., Riesmeier, J.W., Martiny, A., Ramalingam, S., Willmitzer, L. & Trethewey, R.N. (2000) Consequences of the expression of a bacterial glucokinase in potato tubers, both in combination with and independently of a yeast-derived invertase. *Australian Journal of Plant Physiology*, **27**, 827–833.

Fisher, R.A. (1947) The theory of linkage in polysomic inheritance. *Philosophical Transactions of the Royal Society of London*, **B233**, 55–87.

Fregene, M., Angel, F., Gomez, R., *et al.* (1997) A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, **95**, 431–441.

Fry, W.E. & Goodwin, S.B. (1997a) Re-emergence of potato and tomato late blight in the United States. *Plant Disease*, **81**, 1349–1357.

Fry, W.E. & Goodwin, S.B. (1997b) Resurgence of the Irish potato famine fungus. *Bioscience*, **47**, 363–371.

Fu, H.Y. & Park, W.D. (1995) Sink- and vascular-associated sucrose synthase functions are encoded by different gene classes in potato. *Plant Cell*, **7**, 1369–1385.

Fulton, T.M., Xu, Y., Eannetta, N.T., Ilut, D.C., Wright, M.H. & Tanksley, S.D. (2002) Construction of a genetic linkage map of tomato based on conserved ortholog set (COS) markers in common with *Arabidopsis*. Poster presented at PAMG-X, San Diego, 12–16 January (Re Figure 8.3).

Gao, A.G., Hakimi, S.M., Mittanck, C.A., *et al.* (2000) Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nature Biotechnology*, **18**, 1307–1310.

Gebhardt, C. & Valkonen, J.P.T. (2001) Organization of genes controlling disease resistance in the potato genome. *Annual Review of Phytopathology*, **39**, 79–102.

Gebhardt, C., Ritter, E., Barone, A., *et al.* (1991) RFLP maps of potato and their alignment with the homeologous tomato genome. *Theoretical and Applied Genetics*, **83**, 49–57.

Gebhardt, C., Mugniery, D., Ritter, E., Salamini, F. & Bonnel, E. (1993) Identification of RFLP markers closely linked to the H1 gene conferring resistance to *Globodera rostochiensis* in potato. *Theoretical and Applied Genetics*, **85**, 541–544 (Re Figure 8.3).

Geigenberger, P., Stamme, C., Tjaden, J., *et al.* (2001) Tuber physiology and properties of starch from tubers of transgenic potato plants with altered plastidic adenylate transporter activity. *Plant Physiology*, **125**, 1667–1678.

Ghislain, M., Bonierbale, M.& Nelson, R. (1999) *Gene Technology for Potato in Developing Countries*, pp. 105–140. Springer, Vienna.

Ghislain, M., Trognitz, B., Herrera, M.R., *et al.* (2001) Genetic loci associated with field resistance to late blight in offspring of *Solanum phureja* and *S. tuberosum* grown under short-day conditions. *Theoretical and Applied Genetics*, **103**, 433–442.

Greiner, S., Rausch, T., Sonnewald, U. & Herbers, K. (1999) Ectopic expression of a tobacco invertase inhibitor homolog prevents cold-induced sweetening of potato tubers. *Nature Biotechnology*, **17**, 708–711.

Grube, R.C., Radwanski, E.R. & Jahn, M. (2000) Comparative genetics of disease resistance within the *Solanaceae*. *Genetics*, **155**, 873–887.

Györgyey, J., Vaubert, D., Jimenez-Zurdo, J.I., *et al.* (2000) Analysis of *Medicago truncatula* Nodule Expressed Sequence Tags. *Molecular Plant-Microbe Interactions*, **13**, 62–71.

Hahn, S.K., Osiru, D.S.O., Akoroda, M.O. & Otoo, J.A. (1987) Yam production and its future prospects. *Outlook in Agriculture*, **16**, 105–110

Hajirezaei, M.R., Takahata, Y., Trethewey, R.N., Willmitzer, L. & Sonnewald, U. (2000) Impact of elevated cytosolic and apoplastic invertase activity on carbon metabolism during potato tuber development. *Journal of Experimental Botany*, **51**, 439–445.

Hamalainen, J.H., Watanabe, K.N., Valkonen, J.P.T., *et al.* (1997) Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y. *Theoretical and Applied Genetics*, **94**, 192–197 (Re Figure 8.3).

Hamalainen, J.H., Kekarainen, T., Gebhardt, C., Watanabe, K.N. & Valkonen, J.P.T. (2000) Recessive and dominant genes interfere with the vascular transport of potato virus A in diploid potatoes. *Molecular Plant-Microbe Interactions*, **13**, 402–412 (Re Figure 8.3).

Hammond-Kosack, K.E. & Jones, J.D.G. (1997) Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology*, **48**, 575–607.

Hammond-Kosack, K.E., Harrison, K. & Jones, J.D.G. (1994) Developmentally regulated cell death on expression of the fungal avirulence gene Avr9 in tomato seedlings carrying the disease-resistance gene Cf-9. *Proceedings of the National Academy of Sciences, USA*, **91**, 10445–10449.

He, L., Pokrzywa, R., Yencho, C. & Sosinski, B. (2001) Construction of a sweetpotato bac library and its application for resistance gene isolation (abstract). Plant & Animal Genome IX Conference, San Diego, CA, January 13–17, 2001.

Hehl, R., Faurie, E., Hesselbach, J., *et al.* (1999) TMV resistance gene N homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theoretical and Applied Genetics*, **98**, 379–386 (Re Figure 8.3).

Horton, D.E. & Fano, H. (1985) *Potato Atlas.* International Potato Center (CIP), Lima, Peru (Re Figure 8.3).

Howeler, R.H. (1991) Long term effect of cassava cultivation on soil productivity. *Field Crops Research*, **26**, 1–18.

Huber, S.C. & Huber, J.L. (1992) Role of sucrose-phosphate synthase in sucrose metabolism in leaves. *Plant Physiology*, **99**, 1275–1278.

Huber, S.C. & Huber, J.L.A. (1996) Role and regulation of sucrose-phosphate synthase in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **47**, 431–444.

Hurtado, O., Li, X.Q., Kriegner A., Rossel, G. & Zhang, D.P. (2001) Mapping QTL and underlying functional genes associated with carbohydrate metabolism in sweetpotato using AFLP and Cleaved Amplified Polymorphism Sequences (CAPS) (abstract) In: *Sweetpotato: Food and Health for the Future – International Symposium of Sweetpotato.* November 26–29, 2001, Lima, Peru.

Jacobs, J.M.E., Eck, H.J., van Horsman, K., *et al*. (1996) Mapping of resistance to the potato cyst nematode *Globodera rostochiensis* from the wild potato species *Solanum vernei*. *Molecular Breeding*, **2**, 51–60 (Re Figure 8.3).

James, C. (2001) Global review of commercialized transgenic crops: 2000. ISAAA Briefs No. 23. ISAAA, Ithaca, New York.

Jesch, P. (1996) Growth market of starch. Maize leads the world market by far. *Mais*, **4**, 138–139.

Jones, E.S. & St Clair, D. (1998) Mapping QTL for pathotype non-specific resistance from wild tomato (*Lycopersicon hirsutum*) (abstract). Plant and Animal Genome VI. San Diego, January 18–22, 1998.

Jorge, V., Fregene, M.A., Duque, M.C., Bonierbale, M.W., Tohme, J. & Verdier, V. (2000) Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, **101**, 865–872.

Kacser, H.F. & Burns, J.A. (1973) The control of flux. *Symposium of the Society of Experimental Biology*, **27**, 65–104.

Kadowaki, M., Kubota, F. & Saitou, K. (2001) Effects of exogenous injection of different sugars on leaf photosynthesis, dry matter production and adenosine 5'-diphosphate glucose pyrophosphorylase (AGPase) activity in sweetpotato, *Ipomoea batatas* (Lam.). *Journal of Agronomy and Crop Science*, **186**, 37–41.

Kamoun, S., Dong, S., Huitema, E., Vleeshouwers, V.G.A.A. & Govers, F. (1999) Dissection of nonhost resistance of *Nicotiana* to *P. infestans. Trends in Plant Science*, **4**, 196–200.

Kanazin, V., Marek, L.F. & Shoemaker, R. (1996) Resistance gene analogs are clustered in soybean. *Proceedings of the National Academy of Sciences, USA*, **93**, 11746–11750.

Kays, S. (1985) The physiology of yield of sweetpotato. In: *Sweetpotato Products: A Natural Resources for the Tropics* (ed. J.C. Bouwkamp), pp. 79–132. CRC Press, Boca Raton, FL.

Khalil, H.M. & Henry, B.R. (1997) Feasibility of utilizing sweet potato solids in extrusion cooking. *Food and Science Technology International*, **3**, 171–174.

Khul, J.C., Hanneman, R.E. & Havey, M.J. (2001) Characterization and mapping of *Rpi1*, a late blight resistance locus from diploid (1EBN) Mexican *Solanum pinnatisectum*. *Molecular Genetics and Genomics*, **265**, 977–985 (Re Figure 8.3).

Kimura, T., Otani, M., Noda, T., Ideta, O., Shimada, T. & Saito, A. (2001) Absence of amylose in sweet potato [*Ipomoea batatas* (L.) Lam.] following the introduction of granule-bound starch synthase I cDNA. *Plant Cell Reports*, **20**, 663–666.

Kochhar, S.L. (1998) *Economic Botany in the Tropics.* Macmillan India.

Kortstee, A.J. & Vermeesch, A.M.S. (1996) Expression of *Escherichia coli* branching enzyme in tubers of amylose-free transgenic potato leads to an increased branching degree of the amylopectin. *Plant Journal*, **10**, 83–90.

Krause, K.P., Hill, L., Reimholz, R., Nielsen, T.H., Sonnewald, U. & Stitt, M. (1998) Sucrose metabolism in cold-stored potato tubers with decreased expression of sucrose phosphate synthase. *Plant Cell and Environment*, **21**, 285–299.

Kreike, C.M., Koning, J.R.A., de Vinke, J.H., Ooijen, J.W., van Gebhardt, C. & Stiekema, W.J. (1993) Mapping of loci involved in quantitatively inherited resistance to the potato cyst-nematode *Globodera rostochiensis* pathotype Ro1. *Theoretical and Applied Genetics*, **87**, 464–470 (Re Figure 8.3).

Kreike, C.M., Koning, J.R.A., de Vinke, H.J., Ooijen, J.W. & van Stiekema, W.J. (1994) Quantitatively-inherited resistance to *Globodera pallida* is dominated by one major locus in *Solanum spegazzinii*. *Theoretical and Applied Genetics*, **88**, 764–769 (Re Figure 8.3).

Kreike, C.M., Kok-Westeneng, A.A. & Vinke, J.H. (1996) Mapping of QTLs involved in nematode resistance, tuber yield and root development in *Solanum* sp. *Theoretical and Applied Genetics*, **92**, 463–470 (Re Figure 8.3).

Kriegner, A., Cervantes, J.C., Burg, K., Mwanga, R.O. & Zhang, D.P. (2001) A genetic linkage map of sweetpotato (*Ipomoea batatas* [L.] Lam.) based on AFLP markers. In: *Scientist and Farmer – Partners in Research for the 21st Century*, pp.303–314. International Potato Center, Lima, Peru.

Ku, H.M., Vision, T., Liu, J.P. & Tanksley, S.D. (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences, USA*, **97**, 9121–9126.

Kuipers, A.G.J., Vreem, J.T.M., Meyer, H., Jacobsen, E., Feenstra, W.J. & Visser, R.G.F. (1991) Field evaluation of antisense RNA mediated inhibition of GBSS gene expression in potato. *Euphytica*, **59**, 83–91.

Kuipers, A.G.J., Jacobsen, E. & Visser, R.G.F. (1994a) Formation and deposition of amylose in the potato tuber starch granule are affected by the reduction of granule-bound starch synthase gene expression. *Plant Cell*, **6**, 43–52.

Kuipers, A.G.J., Soppe, W.J.J., Jacobsen, E. & Visser, R.G.F. (1994b) Field evaluation of transgenic potato plants expressing an antisense granule-bound starch synthase gene: increase of the antisense effect during tuber growth. *Plant Molecular Biology*, **26**, 1759–1773.

Lander, E.S., Green, P., Abrahamson, J., *et al.* (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181

Lawson, C., Kaniewski, W., Haley, L., *et al.* (1990) Engineering resistance to mixed virus infection in a commercial potato cultivar: Resistance to potato virus X and potato virus Y in transgenic Russet Burbank. *Bio/Technology*, **8**, 127–134.

Leister, D., Ballvora, A., Salamini, F. & Gebhardt, C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics*, **14**, 421–429.

Leonards-Schippers, C., Gieffers, W., Salamini, F. & Gebhardt, C. (1992) The R1 gene conferring race-specific resistance to *Phytophthora infestans* in potato is located on potato chromosome V. *Molecular and General Genetics*, **233**, 278–283.

Leonards-Schippers, C., Gieffers, W., Schafer-Pregl, R., *et al.* (1994) Quantitative resistance to *Phytophthora infestans* in potato: a case study for QTL mapping in an allogamous plant species. *Genetics*, **137**, 67–77.

Li, H.Q., Sautter, C., Potrykus, I. & Puonti-Kaerlas, J. (1996) Genetic transformation of cassava (*Manihot esculenta* Crantz). *Nature Biotechnology*, **14**, 736–740.

Li, X., van Eck, H.J., van der Voort, J.N.A.M.R., Huigen, D.J., Stam, P. & Jacobsen, E. (1998) Autotetraploids and genetic mapping using common AFLP markers: the R2 allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome 4. *Theoretical and Applied Genetics*, **96**, 1121–1128.

Li, X.Q., Haroon, M., Benavides, J., *et al.* (2001) EST development in sweetpotato from cDNA libraries of pencil and storage roots. International Plant & Animal Genome IX, January 13–17, 2001. San Diego, USA.

Lloyd, J.R., Springer, F., Buleon, A., Muller-Rober, B., Willmitzer, L. & Kossmann, J. (1999) The influence of alterations in ADP-glucose pyrophosphorylase activities on starch structure and composition in potato tubers. *Planta*, **209**, 230–238.

Lorberth, R., Ritte, G., Willmitzer, L. & Kossmann, J. (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nature Biotechnology*, **16**, 473–477.

Malcolmson, J.F. & Black, W. (1966) New R-genes in *Solanum demissum* Lindl. and their complementary races of *Phytophthora infestans* (Mont.) de Bary. *Euphytica*, **15**, 199–203.

Marczewski, W., Flis, B., Zyller, J., Schafer-Pregl, R. & Gebhardt, C. (2001) A major quantitative trait locus for resistance to potato leafroll virus is located in a resistance hotspot on potato chromosome XI and is tightly linked to *N*-gene-like markers. *Molecular Plant-Microbe Interactions*, **14**, 1420–1425 (Re Figure 8.3).

Mentaberry, A., Bravo-Almonacid, F., Cabral, S., *et al.* (1997) Development of transgenic potato cultivars with combined resistance to viral and fungal pathogens. Abstract presented at the Workshop Application of Biotechnology Research in Latin America and the Caribbean held in Rio de Janeiro, Brazil, September 1997.

Meyer, R.C., Milborne, D., Hackett, C.A., *et al.* (1998) Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). *Molecular and General Genetics*, **259**, 150–160.

Michelmore, R. (2000) Genomic approaches to plant disease resistance. *Current Opinion in Plant Biology*, **3**, 125–131.

Mignouna, H.D., Mank, R.A., Ellis, T.H.N., *et al.* (2002) A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. *Theoretical and Applied Genetics*, DOI 10.1007/s00122–002–0911–7. Published online: 19 June 2002.

Mihovilovich, E., Alarcon, L., Perez, A.L. & Bonierbale, M. (2001) Discovery and evaluation of a new source of resistance to PLRV: *Solanum tuberosum* subsp. *andigena*. In: *Scientist to Farmer; Partners in Research for the 21st Century.* pp. 88–94. International Potato Center. Program Report 1999–2000. International Potato Center, Lima, Peru.

Mittler, R. & Rizhsky, L. (2000) Transgene-induced lesion mimic. *Plant Molecular Biology*, **44**, 335–344.

Mohan, M., Nair, S., Bhagwat, A., *et al.* (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, **3**, 87–103.

Muller-Rober, B., Sonnewald, U. & Willmitzer, L. (1992) Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. *EMBO Journal*, **11**, 1229–1238.

Munyikwa, T.R.I., Langeveld, S., Salehuzzaman, S.N.I.M., Jacobsen, E. & Visser, R.G.F. (1997) Cassava starch biosynthesis: new avenues for modifying starch quantity and quality. *Euphytica*, **96**, 65–75.

Munyikwa, T.R.I., Kreuze, J., Fregene, M., Suurs, L., Jacobsen, E. & Visser, R.G.F. (2001) Isolation and characterisation of cDNAs encoding the large and small subunits of ADP-glucose pyrophosphorylase from cassava (*Manihot esculenta* Crantz). *Euphytica*, **120**, 71–83.

Naess, S.K., Bradeen, J.M., Wielgus, S.M., Haberlach, G.T., McGrath, J.M. & Helgeson, J.P. (2000) Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8. *Theoretical and Applied Genetics*, **101**, 697–704.

Oberhagemann, P., Chatot-Balandras, C., Schafer-Pregl, R., *et al.* (1999) A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. *Molecular Breeding*, **5**, 399–415.

Ohlrogge, J. & Benning, C. (2000) Unraveling plant metabolism by EST analysis. *Current Opinion in Plant Biology*, **3**, 224–228.

Okada, Y., Saito, A., Mishiguchi, M., *et al.* (2001) Virus resistance in transgenic sweetpotato [*Ipomoea batatas* L. (Lam)] expressing the coat protein gene of sweetpotato feathery mottle virus. *Theoretical and Applied Genetics*, **103**, 743–751.

Ooms, G., Burrell, M.M., Karp, A., Bevan, M. & Hille, J. (1987) Genetic transformation in two potato cultivars with T-DNA from disarmed *Agrobacterium*. *Theoretical and Applied Genetics*, **73**, 744–750.

Ordoñez, M.E., Forbes, G.A. & Trognitz, B.R. (1997) Resistance to late blight in potato: a putative gene that suppresses R genes and its elicited by specific isolates. *Euphytica*, **95**, 167–172.

Ordoñez, M.E., Forbes, G.A. & Trognitz, B. (1998) Relationship between ineffective R-genes and expansion rate of lesions on potato leaves, caused by *Phytophthora infestans. Plant Pathology*, **47**, 130–136.

Osusky, M., Zhou GuoQing, Osuska, L., Hancock, R.E., Kay, W.W. & Santosh, M. (2000) Transgenic plants expressing cationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens. *Nature Biotechnology*, **18**, 1162–1166.

Otani, M., Mii, M., Handa, T., Kamada, H. & Shimada, T. (1993) Transformation of sweetpotato (*Ipomoea batatas* (L.) Lam.) plants by *Agrobacterium rhizogenes*. *Plant Science*, **94**, 151–159 (Re Figure 8.3).

Paterson, A.H., Lin, Y.-R., Li, Z., *et al*. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**, 1714–1718.

Paterson, A.H., Bowers, J.E., Burow, M.D., *et al*. (2000) Comparative genomics of plant chromosomes. *Plant Cell*, **12**, 1523–1539.

Pflieger, S., Lefebvre, V., Blattes, A., *et al*. (1998) Disease resistance gene analogues as candidates for genes and QTLs involved in pepper/pathogen interactions (abstract). Plant and Animal Genome Conference VI. San Diego, January 18–22, 1998.

Pineda, O., Bonierbale, M.W., Plaisted, R.L., Brodie, B.B. & Tanksley, S.D. (1993) Identification of RFLP markers linked to the H1 gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome*, **36**, 152–156 (Re Figure 8.3).

Plaisted, R.L., Bonierbale, M.W., Yencho, G.C., *et al*. (1993) Potato improvement by traditional breeding and opportunities for new technologies. Paper presented at the Potato Molecular Biology Symposium, San Diego, USA.

Prince, J.P., Pochard, E. & Tanksley, S.D. (1993) Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. *Genome*, **36**, 404–417.

Raemakers, C.J.J.M., Sofiari, E., Taylor, N., Henshaw, G., Jacobsen, E. & Visser, R.G.F. (1996) Production of transgenic cassava (*Manihot esculenta* Crantz) plants by particle bombardment using luciferase activity as selection marker. *Molecular Breeding*, **2**, 339–349.

Reymond, P. (2001) DNA microarrays and plant defense. *Plant Physiology and Biochemistry*, **39**, 313–321.

Richmond, T. & Somerville, S. (2000) Chasing the dream: plant EST microarrays. *Current Opinion in Plant Biology*, **3**, 108–116.

Rickert, A.M., Premstaller, A., Gebhardt, C. & Oefner, P.J. (2002) Genotyping of SNPs in polyploid genome by pyrosequencing. *BioTechniques*, **32**, 592–603.

Ritter, E., Debener, T., Barone, A., Salamini, F. & Gebhardt, C. (1991) RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). *Molecular and General Genetics*, **227**, 81–85 (Re Figure 8.3).

Rivera Peña, A. (1990) Screening of *Solanum tuberosum* dihaploids for field resistance to late blight *Phytophthora infestans* (Mont.) de Bary in the Toluca Valley. In: European Association for Potato Research. EAPR abstracts, pp. 10–11. Triennial Conference of the EAPR, July 8–13, Edinburgh, UK.

Rober, M., Geider, K., Mueller, R.B. & Willmitzer, L. (1996) Synthesis of fructans in tubers of transgenic starch-deficient potato plants does not result in an increased allocation of carbohydrates. *Planta*, **199**, 528–536.

Rouppe van der Voort, J.N.A.M., Van Zandvoort, P., Van Eck, H.J., *et al*. (1997) Allele specificity of comigrating AFLP markers used to align genetic maps from different potato genotypes. *Molecular and General Genetics*, **255**, 438 (Re Figure 8.3).

Rouppe van der Voort, J.N.A.M., Van Eckm H.J., Draaistram, J., Van Zandvoort, P.M., Jacobsen, E. & Bakker, J. (1998) An online catalogue of AFLP markers covering the potato genome. *Molecular Breeding*, **4**, 73–77.

Rouppe van der Voort, J., Vossen, E., van de Bakker, E., *et al*. (2000) Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical and Applied Genetics*, **101**, 1122–1130 (Re Figure 8.3).

Rozen, S. & Skaletsky, H.J. (1996) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Safford, R., Jobling, S.A., Sidebottom, C.M., *et al.* (1998) Consequences of antisense RNA inhibition of starch branching enzyme activity on properties of potato starch. *Carbohydrate Polymers*, **35**, 155–168.

Saitou, K., Araki, T., Agata, W., Kubota, F. & Nakayama, K. (1997) [Expression of sucrose synthase in sweet potato]. *Japanese Journal of Crop Science*, **66**, 624–631.

Salanoubat, M. & Belliard, G. (1989) The steady-state level of potato sucrose synthase mRNA is dependent on wounding, anaerobiosis and sucrose concentration. *Gene*, **84**, 181–185.

Salehuzzaman, S.N.I.M., Jacobsen, E. & Visser, R.G.F. (1993) Isolation and characterization of a cDNA encoding granule-bound starch synthase in cassava (*Manihot esculenta* Crantz) and its antisense expression in potato. *Plant Molecular Biology*, **23**, 947–962.

Salehuzzaman, S.N.I.M., Vincken, J.P., Wal, M., van de Straatman-Engelen, I., Jacobsen, E. & Visser, R.G.F. (1999) Expression of a cassava granule-bound starch synthase gene in the amylose-free potato only partially restores amylose content. *Plant Cell and Environment*, **22**, 1311–1318.

Sandbrink, J.M., Colon, L.T., Wolters, P.J.C.C. & Stiekema, W.J. (2000) Two related genotypes of *Solanum microdontum* carry different segregating alleles for field resistance to *Phytophthora infestans. Molecular Breeding*, **6**, 215–225.

Savageau, M.A. (1969) Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology*, **25**, 365–369.

Schopke, C., Taylor, N., Carcamo, R., *et al.* (1996) Regeneration of transgenic cassava plants (*Manihot esculenta* Crantz) from microbombarded embryogenic suspension cultures. *Nature Biotechnology*, **14**, 731–735.

Scott, G.J., Best, R., Rosegrant, M. & Bokanga, M. (2000) Roots and tubers in the global food system: A vision statement to the year 2020 (including annex). Lima (Peru). International Potato Center (CIP); Centro Internacional de Agricultura Tropical (CIAT); International Food Policy Research Institute (IFPRI); International Institute of Tropical Agriculture (IITA); International Plant Genetic Resources Institute (IPGRI).

Simko, I. (2002) Comparative analysis of quantitative trait loci for foliage resistance to *Phytophthora infestans* in tuber-bearing *Solanum* species. *American Journal of Potato Research*, **79**, 125–132.

Simmonds, N.W. & Wastie, R.L. (1987) Assessment of horizontal resistance to late blight of potatoes. *Annals of Applied Biology*, **111**, 213–221.

Simon, R., Herrera, M.R., Portal, L., Ghislain, M. & Bonierbale, M. (2002) Development of consensus tools for association studies of late blight resistance in potato. Poster presented at PAMG X, San Diego, January 12–16.

Smart, C.D., Restrepo, S., Myers, K.L., *et al.* (2002) Differential gene expression during compatible interactions between tomato and *Phytophthora infestans.* 10th International Congress Molecular Plant-Microbe Interactions. University of Wisconsin-Madison. Madison, Wisconsin, July 10–14 (abstract).

Sowokinos, J.R. (2001a) Allele and isozyme patterns of UDP-glucose pyrophosphorylase as a marker for cold-sweetening resistance in potatoes. *American Journal of Potato Research*, **78**, 57–64.

Sowokinos, J.R. (2001b) Biochemical and molecular control of cold-induced sweetening in potatoes. *American Journal of Potato Research*, **78**, 221–236.

Sowokinos, J.R., Thomas, C. & Burrell, M.M. (1997) Pyrophosphorylases in potato. V. Allelic polymorphism of UDP-glucose pyrophosphorylase in potato cultivars and its association with tuber resistance to sweetening in the cold. *Plant Physiology*, **113**, 511–517.

Sowokinos, J.R., Shock, C.C., Stieber, T.D. & Eldredge, E.P. (2000) Compositional and enzymatic changes associated with the sugar-end defect in Russet Burbank potatoes. *American Journal of Potato Research*, **77**, 47–56.

Spencer, D. and Associates (1997) Cassava in Africa: past, present and future. Prepared for the International Institute of Tropical Agriculture. Dunstan Spencer and Associates, Freetown, Sierra Leone.

Spielman, L.J., Drenth, A., Davidse, L.C., *et al.* (1991) A second world-wide migration and population displacement of *Phytophthora infestans*? *Plant Pathology*, **40**, 422–430.

Stark, J.C. & McCann, I.R. (1992) Optimal allocation of limited water supplies for Russet Burbank potatoes. *American Potato Journal*, **69**, 413–421.

Sterky, F., Regan, S., Karlsson, J., *et al.* (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proceedings of the National Academy of Sciences, USA*, **95**, 13330–13335.

Stockhaus, J., Eckes, P., Blau, A., Schell, A. & Willmitzer, L. (1987) Organ-specific and dosage-dependent expression of a leaf/stem specific gene from potato after tagging and transfer into potato and tobacco plants. *Nucleic Acids Research*, **15**, 3479–3491.

Strittmatter, G., Janssens, J., Opsomer, C. & Botterman, J. (1995) Inhibition of fungal disease development in plants by engineering controlled cell death. *Biotechnology*, **13**, 1085–1089.

Sturm, A., Lienhard, S., Schatt, S. & Hardegger, M. (1999) Tissue-specific expression of two genes for sucrose synthase in carrot (*Daucus carota* L.). *Plant Molecular Biology*, **39**, 349–360.

Sweetlove, L.J., Burrell, M.M. & Rees, T. (1996) Starch metabolism in tubers of transgenic potato (*Solanum tuberosum*) with increased ADPglucose pyrophosphorylase. *Biochemical Journal*, **320**, 493–498.

Tai, G.C.C. & Young, D.A. (1984) Early generation selection for important agronomic characteristics in a potato breeding population. *American Potato Journal*, **61**, 419–434.

Takahata, Y., Noda, T. & Sato, T. (1996) Relationship between acid invertase activity and hexose content in sweet potato storage roots. *Journal of Agriculture and Food Chemistry*, **44**, 2063–2066.

Tanksley, S.D., Ganal, M.W., Prince, J.P., *et al.* (1992) High density molecular linkage maps of the potato and potato genomes. *Genetics*, **132**, 1141–1160.

Tauberger, E., Hoffmann-Benning, S., Fleischer-Notter, H., Willmitzer, L. & Fisahn, J. (1999) Impact of invertase overexpression on cell size, starch granule formation and cell wall properties during tuber development in potatoes with modified carbon allocation patterns. *Journal of Experimental Botany*, **50**, 477–486.

Tobias, D.J., Hirose, T., Ishimaru, K., *et al.* (1999) Elevated sucrose-phosphate synthase activity in source leaves of potato plants transformed with the maize SPS gene. *Plant Production Science*, **2**, 92–99.

Tommiska, T.J., Hamalainen, J.H., Watanabe, K.N. & Valkonen, J.P.T. (1998) Mapping of the gene Nx_{phu} that controls hypersensitive resistance to potato virus X in *Solanum phureja* IvP35. *Theoretical and Applied Genetics*, **96**, 840–843 (Re Figure 8.3).

Tooley, P.W. (1990) Variation in resistance to *Phytophthora infestans* among 21 *Solanum verrucosum* plant introductions. *American Potato Journal*, **67**, 491–498.

Trognitz, B.R., Orrillo, M., Portal, L., *et al.* (2001) Evaluation and analysis of reduction of late blight disease in a diploid potato progeny. *Plant Pathology*, **50**, 281–291.

Trognitz, F., Manosalva, P., Gysin, R., *et al.* (2002) Plant defense genes associated with quantitative resistance to potato late blight in *Solanum phureja* × dihaploid *S. tuberosum* hybrids. *Molecular Plant-Microbe Interactions*, **15**, 587–597.

Trowsdale, J. (2002) The gentle art of gene arrangement: the meaning of gene clusters. *Genome Biology*, **3**, Comment 2002.

Tsubone, M., Kubota, F., Saitou, K. & Kadowaki, M. (2000) Enhancement of tuberous root production and adenosine 5'-diphosphate pyrophosphorylase (AGPase) activity in sweetpotato (*Ipomoea batatas* Lam.) by exogenous injection of sucrose solution. *Journal of Agronomy and Crop Science*, **184**, 181–186.

Turkensteen, L.J. (1993) Durable resistance of potatoes against *Phytophthora infestans.* In: *Durability of Disease Resistance. International Symposium on the Durability of Disease*

Resistance in Plants. Wageningen (The Netherlands), February 24–28, 1992 (eds J. Jacobs & J.E. Parlevliet), pp. 115–124. Kluwer Academic, Dordrecht, The Netherlands.

Uchimiya, H., Kidou, S., Shimazaki, T., *et al.* (1992) Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). *Plant Journal*, **2**, 1005–1009.

Umaerus, V. & Umaerus, M. (1994) Inheritance of resistance to late blight. In: *Potato Genetics* (eds J.E. Bradshaw & G.R. Mackay), pp. 365–402. CAB International, Wallingford, UK.

Usuda, H., Demura, T., Shimogawara, K. & Fukuda, H. (1999) Development of sink capacity of the 'storage root' in a radish cultivar with a high ratio of 'storage root' to shoot. *Plant and Cell Physiology*, **40**, 369–377.

Van der Meer, I.M., Ebskamp, M.J.M., Visser, R.G.F., Weisbeek, P.J. & Smeekens, S.C.M. (1994) Fructan as a new carbohydrate sink in transgenic potato plants. *Plant Cell*, **6**, 561–570.

Van Deynze, A.E., Sorrells, M.E., Park, W.D., *et al.* (1998) Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics*, **97**, 356–369.

Vavilov, N.I. (1992) *Origin and Geography of Cultivated Plants.* Cambridge University Press, Cambridge.

Vision, T.J., Brown, D.G., Shmoys, D.B., Durrett, R.T. & Tanksley, S.D. (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics*, **155**, 407–420.

Visser, R.G.F., Somhorst, I., Kuipers, G.J., Ruys, N.J., Feenstra, W.J. & Jacobsen, E. (1991) Inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs. *Molecular and General Genetics*, **225**, 289–296.

Vleeshouwers, V.G.A.A., van Dooijeweert, W., Govers, F., Kamoun, S. & Colon, L.T. (2000) The hypersensitive response is associated with host and nonhost resistance to *Phytophthora infestans. Planta*, **210**, 853–864.

Walbot, V. & Delseny, M. (2002) Genome studies and molecular genetics: genomics approaches to analyzing and using diversity in flowering plants. Editorial overview. *Current Opinion in Plant Biology*, **5**, 91–93.

Wang, H.L., Lee, P.D., Chen, W.L., Huang, D.J. & Su, J.C. (2000) Osmotic stress-induced changes of sucrose metabolism in cultured sweet potato cells. *Journal of Experimental Botany*, **51**, 1991–1999.

Wang ShuJen, Yeh KaiWun & Tsai ChiaYin (1999) Molecular characterization and expression of starch granule-bound starch synthase in the sink and source tissues of sweet potato. *Physiologia Planta*, **106**, 253–261.

Wang, X., Woo, Y.M., Kim, C.S. & Larkins, B.A. (2001) Quantitative trait locus mapping of loci influencing elongation factor 1alpha content in maize endosperm. *Plant Physiology*, **125**, 1271–1282.

Wang, Y., Nelson, R.L. & Hu, Y. (1998) Genetic analysis of resistance to soybean mosaic virus in four soybean cultivars from China. *Crop Science*, **38**, 922–925.

Wastie, R.L. (1991) Resistance to powdery scab of seedling progenies of *Solanum tuberosum*. *Potato Research*, **34**, 249–252.

White, J.A., Todd, J., Newman, T., *et al.* (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiology*, **124**, 1582–1594.

Wilczynski, G., Kulma, A. & Szopa, J. (1998) The expression of 14-3-3 isoforms in potato is developmentally regulated. *Journal of Plant Physiology*, **153**, 118–126.

Woolfe, J.A. (1992) The contribution of sweetpotato and its products to human diets. In: *Sweetpotato Technology for the 21st Century* (eds W.A. Hill, C.K. Bonsi & P.A. Loretan), pp. 367–380. Tuskegee University, USA.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H. & Tanksley, S.D. (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics*, **83**, 294–300.

Yencho, G.C., Cohen, M.B. & Byrne, P.F. (2000) Applications of tagging and mapping insect resistance loci in plants. *Annual Reviews of Entomology*, **45**, 393–422.

Zhang, D.P., Cipriani, G., Rety, I., Golmirzaie, A., Smit, N. & Michaud, D. (2000) Expression of protease inhibitors in sweetpotato. In: *Recombinant Protease Inhibitors in Plants* (ed. D. Michaud), pp. 167–178. Landes Bioscience, Georgetown, TX.

Zimnoch-Guzowska, E., Marczewski, W., Lebecka, R., *et al*. (2000) QLT analysis of new sources of resistance to *Erwinia carotovora* ssp. a*troseptica* in potato done by AFLP, RFLP, and resistance-gene-like markers. *Crop Science*, **40**, 1156–1167 (Re Figure 8.3).

Zrenner, R., Salanoubat, M., Willmitzer, L. & Sonnewald, U. (1995) Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). *Plant Journal*, **7**, 97–107.

Zrenner, R., Krause, K.P., Apel, P. & Sonnewald, U. (1996) Reduction of the cytosolic fructose-1,6-bisphosphatase in transgenic potato plants limits photosynthetic sucrose biosynthesis with no impact on plant growth and tuber yield. *Plant Journal*, **9**, 671–681.

Index